Téléologie et fonctions en biologie. Une approche non causale des explications téléofonctionnelles

Alberto Molina Pérez

UNIVERSIDAD AUTONOMA DE MADRID

Facultad de filosofía

Programa de Doctorado en Estudios Artísticos, Literarios y de la Cultura

TESIS DOCTORAL

TÉLÉOLOGIE ET FONCTIONS EN BIOLOGIE

Une approche non causale des explications téléofonctionnelles

Autor : ALBERTO MOLINA PÉREZ

Director : JESÚS VEGA ENCABO

TRIBUNAL

Javier Ordóñez Rodríguez (Presidente)
Cristian Saborido Alejandro (Secretario)
Françoise Longy (Vocal)

En Madrid, a 19 de diciembre de 2017

RESUMEN (ESPAÑOL)

Esta tesis versa sobre la teleología y las funciones en biología. En concreto, aborda el problema de la legitimidad científica de las atribuciones y de las explicaciones teleofuncionales en biología. Se inscribe en el marco de un debate cuyos orígenes se remontan por lo menos hasta los años setenta del siglo pasado y que sigue estando vigente hoy en día.

La teleología es problemática por sus vínculos históricos con teorías obsoletas y creencias no científicas como el creacionismo, el vitalismo y el finalismo. Por un lado, se ha interpretado como si implicase la existencia de una intencionalidad inmanente o transcendente en los fenómenos naturales. Por otro, se ha interpretado como una forma de causalidad inválida, es decir como causa final o retrógrada. Sin embargo, algunos autores justifican el uso de las explicaciones teleológicas en biología negando que tengan necesariamente implicaciones inaceptables y alegando que desempeñan un papel importante y tal vez imprescindible. La mayoría considera que las explicaciones teleofuncionales no son más que "explicaciones causales disfrazadas" (Papineau, 2005).

La dimensión del debate que me interesa se centra precisamente en la naturalización de las funciones, esto es, en la manera de reducir, traducir o expresar en términos causales las atribuciones y explicaciones funcionales de modo que puedan encajar en el marco de las ciencias de la naturaleza. Tres enfoques se disputan sobre esta cuestión, cada uno de los cuales cuenta con diferentes versiones y ramificaciones.

De acuerdo con el enfoque etiológico o histórico, la función de un rasgo biológico es aquel efecto por el cual ha sido seleccionado en el marco de la teoría darwiniana de la evolución (Wright, 1973; Millikan, 1984; Neander, 1991; Godfrey-Smith, 1994) o, según otros autores, su capacidad actual para contribuir a la selección o a la *fitness* del organismo (Ruse, 1973; Bigelow & Pargetter, 1987; Walsh, 1996). De acuerdo con el enfoque sistémico, la función es el papel causal que desempeña el rasgo en el sistema al que pertenece o, desde otra perspectiva, su contribución a un objetivo de ese sistema entendido desde la teoría cibernética (Cummins, 1975; Boorse, 1976; Nagel, 1977; Davies, 2001). Otros enfoques pretenden unificar los anteriores o recurren a otras nociones no

menos problemáticas, como las de diseño y de valor (Woodfield, 1976; Bedau, 1992; Kitcher, 1993; Griffiths, 1993; McLaughlin, 2001).

Mi enfoque se desmarca de los anteriores porque no asumo la premisa según la cual las explicaciones teleológicas son explicaciones causales disfrazadas. Al contrario, afirmo que son aceptables *como tales*, tanto de manera general como en el ámbito de las ciencias actuales y en particular en el de la biología.

De forma general, pongo en tela de juicio la idea según la cual el pensamiento teleológico sería un pensamiento causal inmaduro, así como la idea de su dependencia respecto de la psicología, es decir, de la intencionalidad. Rechazo pues las acusaciones de antropomorfismo, de vitalismo y de finalismo que han sido formuladas contra ella. Por lo contrario, defiendo que la teleología, la causalidad y la intencionalidad se corresponden con modos diferentes, autónomos y complementarios de representación del mundo exterior.

En el ámbito científico, pongo de manifiesto que las explicaciones causales y teleológicas son diferentes y complementarias, y que tanto unas y otras pueden ser sometidas a procesos de escrutinio racional para determinar si son falsas o inválidas. Muestro también que la teoría de la explicación de Woodward (1997, 2003) puede ser generalizada a las explicaciones teleológicas y que estas últimas aportan una contribución original a la explicación científica de los fenómenos naturales en la medida en que permiten definir nuevos invariantes.

En biología, la teleología permite formular predicciones que serían difíciles o imposibles de realizar de otro modo, como la aparición de organismos capaces de alimentarse de plástico o las características de los roedores eusociales. Desempeña también un papel muy importante en la clasificación de los seres vivos y hace posible numerosas inferencias y generalizaciones.

Mi justificación de la teleología se basa en parte en el principio de acción racional o eficiente. Se trata de un principio de optimización extraído de la teoría de la actitud teleológica desarrollada en psicología cognitiva (Gergely y Csibra, 2003; Liu & Spelke, 2017). Esta teoría se basa a su vez en la de Dennett (1987), aunque introduce una estrategia predictiva nueva y diferente de las estrategias intencional (*intentional stance*) y del diseño (*design stance*). El principio de acción eficiente implica que los medios (la acción), los fines (los resultados previstos) y las constricciones o condicionantes (el contexto) mantienen una relación invariante, de modo que cada uno de esos elementos se puede inferir a partir de los demás. No es pues un principio ontológico ni metafísico, sino epistemológico. Impone restricciones en cuanto a las explicaciones teleológicas que se pueden dar por aceptables y permite por otra parte la formulación de predicciones empíricas contrastables.

Al igual que la mayoría de los autores implicados en el debate, propongo también una definición del concepto de función. Ésta viene a decir que las funciones son relaciones de medios a fines en el seno de una totalidad organizada y en un contexto determinado. Las atribuciones funcionales aportan a la vez una explicación no causal de la presencia de un rasgo en un sistema determinado y una explicación del papel causal de ese rasgo en el sistema como medio de un fin. Se trata de una concepción suficientemente general y abstracta como para unir las concepciones etiológicas y sistémicas, y es efectivamente compatible con varias de ellas entendidas como casos particulares de un caso más general. Por otra parte, es aplicable por igual a las funciones biológicas y técnicas.

El método sobre el que se apoya esta definición es la *explicación filosófica* propuesta por Carnap (1950) y por Quine (1960) antes de ser retomada por Schwartz (2004). Se distingue del análisis conceptual empleado por muchos de los autores implicados en el debate, pues no trata de capturar el significado del concepto, ni de describir lo que la gente tiene en la mente cuando lo emplea, ni de desvelar sus condiciones de aplicación. Consiste más bien en mostrar cómo un concepto ambiguo y problemático puede ser sustituido por otro más preciso que desempeña el mismo papel, pero que no plantea los mismos problemas.

Este trabajo consta de catorce capítulos divididos en cinco partes. La primera se centra en el debate antes mencionado sobre la naturalización de las funciones. Examinamos tres tipos de enfoques. El etiológico reune aquellas posturas que se apoyan en la dimensión temporal o histórica de las funciones, especialmente en el marco de la teoría darwiniana de la evolución (CAP. I). El sistémico reune aquellas que sitúan las funciones en el marco de la actividad de un sistema, bien sea en términos de papel causal o de contribución a un fin (CAP. II). Los enfoques mixtos hacen hincapié en algunas de las intuiciones de cada uno de los dos anteriores e intentan unificarlas bajo una concepción única (CAP. III).

La segunda parte examina el problema de las funciones desde la perspectiva de las investigaciones científicas sobre el fenómeno de la vida, es decir desde aquellas disciplinas que estudian las características mínimas de los seres vivos, como la biología sintética, con el objetivo de intentar explicar o caracterizar lo que son las funciones (CAP. IV). Los problemas filosóficos que plantean estas investigaciones científicas nos llevan a poner en cuestión la realidad y la objetividad de la vida y de sus fronteras. En concreto, señalamos que si la distinción entre la vida y la no-vida fuera relativa al observador, entonces la misma conclusión sería aplicable a las funciones biológicas (CAP. V).

En la tercera parte, propongo una nueva concepción de las funciones biológicas y técnicas que pretende unificar los enfoques etiológicos y sistémicos, y superar el reduccionismo causal. En primer lugar, explico en qué consiste y discuto algunas de sus implicaciones (CAP. VI).

A continuación, propongo una formulación precisa de esta concepción y aporto algunos elementos de justificación (CAP. VII).

La cuarta parte está dedicada a justificar la legitimidad científica de mi concepción no causal de la finalidad. Primero, desde la perspectiva de la psicología cognitiva, destaco que el modo de pensar teleológico es probablemente innato y autónomo respecto de los modos de pensamiento causal e intencional (CAP. VIII). Luego, partiendo de la discusión de casos muy sencillos, comparo las explicaciones causales y teleológicas para mostrar en qué consisten y en qué se diferencian (CAP. IX). Por otra parte, rechazo algunas de las objeciones y acusaciones habitualmente lanzadas contra la teleología, como la causalidad retrógrada, el finalismo, el vitalismo y el antropomorfismo (CAP. X). Por último, pongo de manifiesto que las atribuciones y las explicaciones teleofuncionales en biología desempeñan un papel científico muy importante y complementario de las explicaciones causales (CAP. XI).

La quinta y última parte se centra en tres enfoques alternativos desde los cuales es posible también justificar la teleología en biología. El enfoque mentalista (CAP. XII) dice que la atribución de fines a los seres vivos y a las máquinas descansa en una metáfora o una analogía con la intencionalidad humana (Ducasse, 1925,1959; Nissen, 1997; Woodfield, 1976; Dennett, 1987,1991). El enfoque valorativo (CAP. XIII) dice que las atribuciones de fines y de funciones implican atribuciones de valores, y que estas últimas sí están justificadas en biología (Woodfield, 1976; Searle, 1995; Bedau 1992; McLaughlin, 2001). El enfoque organizacional (CAP. XIV) es una prolongación reciente de los enfoques mixtos que pretenden unificar las concepciones etiológicas y sistémicas, pero se distingue de ellos por restringir las atribuciones funcionales a aquellos sistemas que son capaces de automantenerse o de auto-(re)producirse (Schlosser, 1998; McLaughlin, 2001; Christensen & Bickhard, 2002; Mossio, Saborido & Moreno, 2009; Toepfer, 2012). El análisis de estos tres enfoques es una oportunidad para precisar y matizar comparativamente mi propia postura, y me permite también abordar con más detalle la cuestión de la objetividad de las funciones.

Este trabajo aporta varias contribuciones originales al debate sobre las funciones. En primer lugar, analiza el concepto desde una perspectiva interdisciplinar que incluye los debates en filosofía de la biología y de la técnica, en biología sintética y en psicología cognitiva. En segundo lugar, propone una concepción general del concepto de función que permite unificar muchas otras concepciones. En tercer lugar, señala la existencia de una solución radicalmente diferente para el problema de la legitimidad de las atribuciones y de las explicaciones teleológicas en biología que no pasa por su reducción o su traducción en términos no-causales, sino por la búsqueda de sus propias condiciones de validez científica.

RÉSUMÉ (FRANÇAIS)

Ce travail de thèse porte sur la téléologie et les fonctions en biologie. Plus précisément, il porte sur la légitimité scientifique des attributions et des explications téléofonctionnelles en biologie. Il s'inscrit dans le cadre d'un débat à plusieurs facettes que l'on peut faire remonter au moins jusqu'aux années 1970 et qui est encore très actif aujourd'hui.

La téléologie en biologie est problématique parce qu'elle évoque des théories caduques et des croyances non-scientifiques, comme le vitalisme, le finalisme et le créationnisme. Les explications téléologiques semblent en effet impliquer soit une intentionnalité à l'œuvre dans les phénomènes naturels, immanente ou transcendante, soit une forme invalide de la causalité, par exemple sous la forme de causes finales ou rétrogrades. Nonobstant, certains auteurs essayent de les justifier en montrant qu'elles n'ont pas nécessairement d'implications scientifiques inacceptables et qu'elles jouent en biologie un rôle très important, voire irremplaçable. Parmi eux, la plupart estiment que les explications téléofonctionnelles sont en réalité des « explications causales déguisées » (Papineau, 2005).

Par conséquent, l'un des aspects du débat porte sur la naturalisation des fonctions, c'est-à-dire sur la manière de les réduire, de les traduire ou de les expliciter en termes de causes efficientes de sorte qu'elles trouvent leur place dans le cadre des sciences de la nature. Plusieurs approches sont en lice, comportant chacune plusieurs versions et ramifications.

L'approche étiologique consiste à dire que la fonction d'un trait est soit l'effet pour lequel ce trait a été historiquement sélectionné dans le cadre de la théorie darwinienne de l'évolution (Wright, 1973; Millikan, 1984; Neander, 1991; Godfrey-Smith, 1994), soit sa disposition actuelle à contribuer à la sélection ou à la *fitness* de l'organisme (Ruse, 1973; Bigelow & Pargetter, 1987; Walsh, 1996). L'approche systémique consiste à dire que la fonction d'un trait est son rôle causal dans le système auquel il appartient ou sa contribution à un but entendu en termes cybernétiques (Cummins, 1975; Boorse, 1976; Nagel, 1977; Davies, 2001). D'autres approches cherchent à unifier les précédentes ou font intervenir des notions problématiques, comme celles de *design* et de valeur (Woodfield, 1976; Bedau, 1992; Kitcher, 1993; Griffiths, 1993; McLaughlin, 2001).

Notre approche ici est radicalement différente, car nous remettons en question la prémisse selon laquelle les explications téléologiques seraient des explications causales déguisées. Nous défendons au contraire qu'elles sont acceptables *en tant que telles* aussi bien de façon générale que dans le domaine scientifique et en particulier en biologie.

De façon générale, nous remettons en question l'idée selon laquelle la pensée téléologique serait une pensée causale immature, ainsi que sa dépendance présumée vis-à-vis de la psychologie, c'est-à-dire de l'intentionnalité. Nous rejettons donc les accusations d'anthropomorphisme, de vitalisme et de finalisme formulées contre elle. Nous défendons au contraire que la téléologie, la causalité et l'intentionnalité correspondent à des modes différents, autonomes et complémentaires de représentation du monde extérieur.

Dans le domaine scientifique, nous montrons que les explications causales et téléologiques sont différentes et complémentaires, et que les unes et les autres peuvent être invalidées et falsifiées. Nous montrons aussi que la théorie de l'explication causale de Woodward (1997, 2003) peut être généralisée aux explications téléologiques, et que ces dernières apportent une contribution irremplaçable à l'explication scientifique des phénomènes naturels dans la mesure où elles permettent d'identifier un type particulier d'invariants.

En biologie, la téléologie permet de formuler des prédictions qui seraient difficiles ou impossibles à réaliser autrement, comme l'apparition d'organismes mangeurs de plastique ou les caractéristiques des rongeurs eusociaux. Elle joue aussi un rôle très important dans la classification des êtres vivants et rend possible de nombreuses inférences et généralisations.

Notre justification de la téléologie s'appuie en partie sur le principe d'action rationnelle ou efficiente, qui est un principe d'optimalité tiré de la théorie de l'attitude téléologique en psychologie cognitive (Gergely & Csibra, 2003; Liu & Spelke, 2017). Cette théorie s'inspire à son tour de celle de Dennett (1987), mais elle y introduit une stratégie prédictive nouvelle et différente des stratégies intentionnelle (*intentional stance*) et artefactuelle (*design stance*). Le principe d'action efficiente implique que les moyens (l'action), les fins (les résultats attendus) et les contraintes de la situation (le contexte) entretiennent une relation invariante, de sorte que l'on peut inférer l'un des éléments à partir des deux autres. Ce n'est donc pas un principe ontologique ni métaphysique, mais épistémologique. Il impose en effet des restrictions quant aux explications téléologiques que l'on peut considérer comme acceptables et il permet par ailleurs la formulation de prédictions empiriques falsifiables.

À l'instar de la plupart des auteurs impliqués dans le débat, nous avançons aussi une définition du concept de fonction. Elle consiste à dire que les fonctions sont des relations moyen–fin au sein d'un tout organisé et dans un contexte donné. Les attributions fonctionnelles apportent à la fois une explication non-causale de la présence d'un trait dans un système

donné et une explication du rôle causal de ce trait au sein du système en tant que moyen d'une fin. Cette conception est suffisamment générale et abstraite pour unifier les conceptions étiologiques et systémiques, et on peut montrer qu'elle est effectivement compatible avec plusieurs d'entre elles entendues comme des cas particuliers d'un cas plus général. Elle est par ailleurs applicable à la fois des fonctions biologiques et techniques.

La méthode sur laquelle s'appuie cette définition est l'*explication philosophique* proposée par Carnap (1950) et par Quine (1960), puis reprise par Schwartz (2004). Contrairement à l'analyse conceptuelle employée par de nombreux auteurs impliqués dans le débat, il ne s'agit pas de capturer la signification du concept, ni de décrire ce que les gens ont à l'esprit quand ils l'emploient, ni de rendre manifestes ses conditions d'application. Il s'agit plutôt de montrer qu'un concept vague ou problématique peut être avantageusement remplacé par un autre plus précis qui joue le même rôle, mais qui ne pose pas les mêmes problèmes.

Ce travail se compose de quatorze chapitres divisés en cinq parties. La première est consacrée au débat sur la naturalisation des fonctions. Nous y examinons trois types d'approches. L'approche étiologique regroupe les postures qui accordent aux fonctions une dimension temporelle ou historique, notamment dans le cadre de la théorie darwinienne de l'évolution (CHAP. I). L'approche systémique regroupe les postures qui situent les fonctions dans le cadre de l'activité d'un système, que ce soit en termes de rôle causal ou de contribution à un but (CHAP. II). Les approches mixtes retiennent certaines intuitions de chacune des deux approches précédentes pour essayer de les unifier sous une conception unique (CHAP. III).

La deuxième partie est consacrée à l'examen des fonctions depuis la perspective des recherches sur le vivant. Nous nous y intéressons à la biologie synthétique et aux recherches sur les origines de la vie, c'est-à-dire aux disciplines qui étudient les caractéristiques minimales des êtres vivants, pour essayer d'expliquer ou de caractériser, d'un point de vue scientifique, ce que sont les fonctions (CHAP. IV). Les problèmes philosophiques que soulèvent ces recherches nous amènent à nous interroger sur la réalité et sur l'objectivité du vivant et de ses frontières. Nous signalons notamment que si la distinction entre le vivant et l'inerte dépendait en quelque sorte de l'observateur, alors il pourrait en aller de même des fonctions biologiques (CHAP. V).

La troisième partie est consacrée à la formulation d'une nouvelle conception des fonctions biologiques et techniques, afin d'unifier les approches étiologique et systémique, et de dépasser le réductionnisme causal. Nous commençons par expliquer en quoi elle consiste et par discuter certaines de ses implications (CHAP. VI). Nous proposons ensuite une formulation précise de cette conception et apportons quelques éléments de justification (CHAP. VII).

La quatrième partie est consacrée à la justification du recours à une conception non-causale de la finalité et à sa légitimité épistémologique. Nous montrons d'abord, du point de vue de la psychologie cognitive, que le mode de raisonnement téléologique est vraisemblablement inné et autonome vis-à-vis des modes de raisonnement causal et intentionnel (CHAP. VIII). Ensuite, à partir d'un ensemble de cas très simples, nous comparons les explications causales et téléologiques pour montrer en quoi elles consistent et en quoi elles diffèrent (CHAP. IX). Par ailleurs, nous rejetons certaines des objections et des accusations récurrentes formulées contre la téléologie, comme la causalité rétrograde, le finalisme, le vitalisme et l'anthropomorphisme (CHAP. X). Finalement, nous montrons que les attributions et les explications téléofonctionnelles en biologie jouent un rôle scientifique très important et complémentaire de celui des explications causales (CHAP. XI).

La cinquième partie est consacrée à trois approches alternatives à partir desquelles il est également possible de justifier l'attribution de fins à des objets naturels. L'approche mentaliste (CHAP. XII) consiste à dire que l'attribution d'une finalité aux êtres vivants et aux machines relève essentiellement de la métaphore ou de l'analogie avec l'intentionnalité humaine (Ducasse, 1925, 1959; Nissen, 1997; Woodfield, 1976; Dennett, 1987, 1991). L'approche valorative (CHAP. XIII) consiste à dire que les attributions de fins et de fonctions impliquent des attributions de valeurs, et que ces dernières sont justifiées (Woodfield, 1976; Searle, 1995; Bedau 1992; McLaughlin, 2001). L'approche organisationnelle (CHAP. XIV) est un prolongement récent des approches mixtes qui cherchent à unifier les conceptions étiologiques et systémiques, mais elle se distingue par la restriction des attributions fonctionnelles aux systèmes capables de s'auto-maintenir ou de s'auto-(re)produire (Schlosser, 1998; McLaughlin, 2001; Christensen & Bickhard, 2002; Mossio, Saborido & Moreno, 2009; Toepfer, 2012). L'analyse de ces trois approches nous permet de préciser et de nuancer par comparaison notre propre posture, et elle nous permet d'aborder plus en détail la question de l'objectivité des fonctions.

Ce travail apporte plusieurs contributions originales à la réflexion sur les fonctions biologiques. En premier lieu, il analyse le concept depuis une perspective interdisciplinaire qui inclut les débats en philosophie de la biologie et de la technique, en biologie synthétique et en psychologie cognitive. En second lieu, il propose une conception générale du concept de fonction qui permet d'unifier un certain nombre d'autres conceptions. En troisième lieu, il suggère une solution radicalement différente au problème de la légitimité des attributions et des explications téléologiques en biologie qui ne passe pas par leur réduction ni leur traduction en termes de causes, mais par l'investigation de leurs conditions de validité.

AGRADECIMIENTOS

Para mi arcángel, que queda exento de leer este ~~peñazo~~ sesudo trabajo*.

Uno no escribe una tesis doctoral sin el apoyo de muchas personas humanas y no-humanas a las que quiero aquí dar las gracias. En primer lugar, a Kant por prestarme dos o tres ideas y no pedirme nunca que se las devolviera. En segundo lugar, perro no menos importante, a Otto por todos los lametones que me dió y que tampoco le devolví.

Antes de las funciones biológicas, el tema al que quise dedicar la tesis era un poquitiño más extenso, algo así como *la vida, el universo y todo lo demás*. Tuve la suerte de que Michel Bitbol me aceptara como doctorando. También tuve la suerte de poder seguir las clases de filosofía de la biología impartidas por Jean Gayon en el IHPST, las cuales fueron el núcleo de condensación de la investigación desarrollada en este trabajo. Durante mis años en París tuve además la oportunidad de trabajar con Eugenia Lamas con quien aprendí de bioética y de muchas otras cosas.

Doy gracias también a aquellos profesores de Salamanca que son un ejemplo en lo intelectual como en lo personal y que me han regalado su amistad, especialmente Luciano Espinoza y Teresa López de la Vieja que siempre estuvo ahí para ayudarme. En fecha más reciente descubrí los tesoros humanos de Granada, entre los cuales están Arancha San Ginés, Javier Rodríguez Alcázar y Lilian Bermejo-Luque que me ayudaron a mejorar partes de la tesis. Pero sobre todo quiero agradecer a Jesús Vega por su confianza, su apoyo, su dedicación y su generosidad, sin los cuales probablemente no habría terminado la tesis.

Gracias a mi familia por estar ahí y poder contar con ellos, y me refiero aquí no sólo a mis padres y hermanos, sino también a los de Cris que me han adoptado, así como a dos hermanos no de sangre pero sí de alma y corazón, Manouk y David, que me han empujado por activa y por pasiva a terminar de una vez. Por último, gracias a Cris por su paciencia, su sacrificio, su inspiración y su amor.

* Se lo contará su padre de viva voz.

SOMMAIRE

Il est bien difficile au philosophe de s'exercer à la philosophie biologique sans risquer de compromettre les biologistes qu'il utilise ou qu'il cite. Une biologie utilisée par un philosophe n'est-ce pas déjà une biologie philosophique, donc fantaisiste ? Mais serait-il possible, sans la rendre suspecte, de demander à la biologie l'occasion, sinon la permission, de repenser ou de rectifier des concepts philosophiques fondamentaux, tels que celui de vie ?

Georges Canguilhem, *La Connaissance de la vie* (1965, p. 83)

INTRODUCTION GÉNÉRALE

> Longtemps le biologiste s'est trouvé devant la téléologie comme auprès d'une femme dont il ne peut se passer, mais en compagnie de qui il ne veut pas être vu en public.
>
> François Jacob, *La Logique du vivant* (1970, p. 17)

L'objectivité de la biologie pose un problème ontologique et un problème épistémologique. D'un côté, nous ne savons toujours pas quel est (ou ce qu'est) cet objet, ce $\beta\iota o\varsigma$, auquel la biologie se rapporte en dernière instance. De l'autre, la légitimité scientifique des explications téléologiques et fonctionnelles est depuis longtemps une question sujette à controverses. Ces problèmes sont doublés de difficultés conceptuelles qui concernent respectivement les termes de « vie » et de « fonction ».

Or, la vie et les fonctions semblent être intimement liées. En effet, nous n'attribuons de fonctions qu'à ce qui relève ou qui appartient au domaine du vivant, c'est-à-dire aux organes, structures et processus des êtres vivants, à leurs actions et comportements, aussi bien individuels que collectifs, à leurs émotions, pensées et processus cognitifs, aux produits de ces actions et pensées, comme les artefacts et leurs parties, le langage, etc.

Dans ce contexte il semble bien difficile d'examiner l'un des deux concepts sans se pencher aussi sur l'autre. En explorant la question des fonctions et de la téléologie en biologie, nous seront donc amenés à nous interroger sur les liens qui les lient au vivant et à aborder simultanément, ou presque, les problèmes ontologique et épistémologique de la biologie, bien que notre attention sera centrée principalement sur le second.

L'absence d'une définition rigoureuse de la vie n'a pas empêché la biologie de se développer en tant que science rigoureuse. Elle n'a pas même été perçue comme un problème par beaucoup de biologistes qui ont refusé de se poser la question ou qui ont refusé d'y répondre. Et puisque d'autres sciences, comme l'astronomie et la physique, se trouvent dans une situation similaire quant à leurs concepts de base, on peut se

demander si de telles définitions sont toujours nécessaires, voire utiles. Quoi qu'il en soit, cette absence était compensée jusqu'à présent par l'identification relativement aisée des êtres vivants par opposition aux objets inanimés. Pourtant, depuis quelques années, les progrès de la biologie synthétique et de l'exploration spatiale ont rendu problématique leur identification, car de nouvelles entités aspirent au statut d'êtres vivants tout en repoussant les limites de ce que nous reconnaissons habituellement comme tel. Il s'agit de la vie artificielle (organique ou *in silico*), des formes de vie extraterrestre, et des objets situés aux origines de la vie, à la frontière entre l'inerte et le vivant. Faut-il étendre les frontières de la biologie jusqu'aux artefacts moléculaires, aux programmes informatiques et aux organismes exotiques que nous rencontrerons peut-être sur d'autres planètes ? En reposant le problème de la nature du vivant, ces entités nouvelles remettent en question les limites et l'unité des sciences de la vie. Elles déroutent nos habitudes de pensée, troublent nos intuitions catégorielles et dérangent l'ameublement ontologique de l'univers. Au-delà des communautés scientifiques directement impliquées (exobiologistes, informaticiens, biochimistes…), nous sommes tous concernés par le statut de ces objets dont l'apparition affectera notre représentation du monde et transformera nos modes de vie.

D'après Jacques Monod (1970), les êtres vivants se distinguent de tous les autres objets présents dans l'univers par le fait qu'ils sont doués de ce qu'il appelle un « projet » ; et le problème central de la biologie, selon lui, est la contradiction entre cette propriété des êtres vivants et le postulat d'objectivité de la nature qui nie la validité des explications formulées en termes de fins. Les explications téléologiques sont en effet souvent associées à des doctrines métaphysiques pré-darwiniennes comme le vitalisme, l'animisme et le créationnisme ; elles sont aussi perçues comme impliquant un finalisme à la Teilhard de Chardin ou encore une causalité rétrograde (du futur vers le présent) qui les rend incompatibles avec les explications mécanistes des sciences de la nature. La biologie actuelle continue néanmoins d'y avoir recours, quoique de manière souvent implicite et parfois honteuse comme nous le rappelle la citation de François Jacob en épigraphe, attribuée à J.B.S. Haldane.

Au-delà de l'explication scientifique de la téléologie ou téléonomie que proposent des auteurs comme Monod, la question philosophique est de savoir si les notions de fin et de fonction peuvent être éliminées du langage de la biologie ou si, contrairement à ce qui se passe dans les sciences physico-chimiques, elles y jouent un rôle essentiel dont il faudrait rendre compte. Cette question nourrit un important débat dans la littérature spécialisée depuis la seconde moitié du XX^{e} s., et notamment depuis les années 1970. C'est à lui que nous allons consacrer l'essentiel de ce travail en essayant d'y apporter notre contribution.

Aperçu des débats

L'origine du débat est en partie liée à l'analyse du discours et des méthodes de la science et, en particulier, à l'analyse de ses explications. Ainsi, dès la première moitié du XX^e^ s., certains proposaient de définir précisément les notions d'« explication », de « finalité » (*purposiveness*) et d'« explication en termes de fin » pour mieux distinguer les explications mécanique et téléologique (Ducasse, 1925).

Vingt ans plus tard, dans un article précurseur de la cybernétique, Arturo Rosenblueth, Norbert Wiener et Julian Bigelow (1943) annonçaient une nouvelle méthode scientifique pour étudier tous types de phénomènes. Ils y revendiquaient les concepts de finalité (*purpose*) et de téléologie qui, bien que discrédités, leurs semblaient importants. Une discussion s'ensuivit avec Richard Taylor à propos de leur rôle dans les sciences et de leur rapport à l'intentionnalité (Rosenblueth & Wiener, 1950; Taylor, 1950a, 1950b). Tandis que les premiers défendaient la téléologie comme étant une approche scientifique légitime pour analyser le comportement de certains systèmes naturels et mécaniques, Taylor excluait le recours à la finalité dans les explications mécaniques et définissait cette notion en termes de désirs et de croyances, c'est-à-dire qu'il la réservait pour expliquer le comportement des agents intentionnels.

À la même époque, Richard B. Braithwaite (1946) distinguait les explications téléologiques des « explications causales ordinaires », les premières faisant selon lui référence à des événements futurs, ce qui les rendait problématiques dans le cas de la biologie. Une première solution, disait-il, consiste à souligner l'analogie entre les explications téléologiques de la biologie et celles portant sur les actions intentionnelles, et les réduire toutes à des explications où une intention ou quelque chose d'analogue à une intention est la cause efficiente. La seconde consiste à les réduire à des explications physico-chimiques ordinaires. Après les avoir rejetées toutes deux, il proposait la sienne — qui sera ensuite formulée en termes de plasticité et de persistance d'un comportement (ou chaîne causale) vers un but (ou état final).

À partir des années 1950 et '60, de nombreux auteurs prirent part à la discussion (parmi lesquels Beckner, 1959, 1969, Bertalanffy, 1950, 1968, Canfield, 1964, 1966; J. Cohen, 1951; Ducasse, 1959; Frankfurt & Poole, 1966; Gruner, 1966; Harris, 1959; Hempel, 1965; Lehman, 1965a, 1965b; Mayr, 1961; Nagel, 1953, 1961; Scheffler, 1959; Sommerhoff, 1950; Sorabji, 1964; Williams, 1966; L. Wright, 1968). L'enjeu principal était de déterminer si les explications téléofonctionnelles étaient scientifiquement acceptables et comment il fallait les comprendre ou les interpréter pour que ce fût le cas.

Pour ceux qui les jugeaient acceptables, il s'agissait de montrer qu'elles étaient compatibles avec les explications causales ordinaires, c'est-à-dire avec celles de la mécanique et de la physicochimie, ou avec celles se

rapportant aux actions intentionnelles (suivant Davidson, 1963), ou bien encore de montrer l'existence d'une troisième voie. Autrement dit, il s'agissait pour beaucoup de montrer que le discours téléologique, et en particulier celui que l'on trouve dans les sciences de la vie, est équivalent à d'autres formes de discours scientifique qui, elles, sont certainement respectables.

Cela dit, cette équivalence pouvait être interprétée de plusieurs manières. Pour certains, il s'agissait d'une équivalence logique, c'est-à-dire que les explications téléologiques et non-téléologiques avaient le même sens :

> « The position that every legitimate teleological description is, or must be, equivalent to some non-teleological description of the same phenomenon has been expressed in several different ways. The teleological account has been said to be 'translatable into', 'reducible to', or merely to be 'saying the same thing as' the non-teleological account. More elaborately, the point has been made by saying that the two accounts have the same 'cognitive content' and differ only in 'emphasis'. The equivalence between the two kinds of accounts being asserted by these expressions is a logical equivalence of some sort. Expressions like 'same cognitive content', 'translatable', 'saying the same thing' all concern the sense of the descriptions. Accordingly, the teleological description is to have the same sense as the non-teleological one. The two are logically equivalent descriptions. » (L. Wright, 1968, p. 211-2)

Pour d'autres, cela voulait dire que les explications téléologiques n'étaient pas moins scientifiquement valables que les non-téléologiques, quoique différentes. Car bien que tous les phénomènes spatio-temporels pussent être décrits d'une façon comme de l'autre, les deux descriptions n'étaient pas logiquement équivalentes ni ne disaient les mêmes choses :

> « [S]ince every teleological description is of a phenomenon taking place, as it were, in space-time, it will have a corresponding non-teleological description of the physico-geometrical sort. And these two descriptions will not be logically equivalent because they are saying something different about what is taking place. This point is often put by saying that the teleological account says more than the physico-geometrical account. The teleological account says that there is such-and-such a series of events, but then says something else in addition. » (L. Wright, 1968, p. 213)

Derrière la question de l'équivalence des discours, il y avait des enjeux métaphysiques et épistémologiques importants. L'un d'eux était l'autonomie de la biologie. La question de l'unité des sciences a en effet préoccupé les philosophes pendant une bonne partie du XX^e^ siècle, et en particulier entre les années 1925 et 1970. Or, la thèse minimale de l'unité des sciences reposait sur deux idées : « à savoir qu'il y a un ensemble

unique de phénomènes qui constitue l'objet unique de toutes les sciences, et que tous les énoncés scientifiques sont exprimés dans un langage unique (ou sont au moins en principe traduisibles en un langage unique) » (Kistler, 2013).

Par conséquent, il s'agissait de savoir, d'abord, si les phénomènes décrits par la biologie et ceux décrits par la physique appartenaient au même ensemble unique qui est l'objet de toutes les sciences, ou si, au contraire, il existait une hétérogénéité radicale entre les phénomènes vivants et inertes. En toile de fond de cette question se trouvait le débat ontologique entre le mécanisme et le vitalisme. Ce dernier, bien que s'étant éteint dans la première moitié du XX^e^ siècle, allait rester dans les esprits et servir d'épouvantail longtemps après.

Ensuite, il s'agissait de savoir si les énoncés de la biologie et, en particulier, les attributions et les explications téléofonctionnelles, pouvaient être réduites ou traduites, ou encore expliquées en des termes compatibles avec ceux des autres sciences. Car même en admettant que les êtres vivants sont composés de matière inerte et que toutes leurs propriétés sont déterminées par celles de leurs parties et par les relations entre elles, et même en acceptant le physicalisme, il n'en demeure pas moins que la biologie peut constituer une discipline autonome dont les explications et les méthodes sont essentiellement différentes et irréductibles à celles des autres sciences de la nature.

C'est la question que Ernest Nagel examinait dans *The Structure of Science* pour défendre l'unité des sciences et la réductibilité de la biologie à la physicochimie :

> « Most biologists are in general agreement that vital processes, like nonliving ones, occur only under determinate physicochemical conditions and form no exceptions to physicochemical laws. Some of them nevertheless maintain that the mode of analysis required for understanding living phenomena is fundamentally different from that which obtains in the physical sciences. […] In any event, it is instructive to examine some of the reasons biologists commonly advance for the claim that the logic of explanatory concepts in biology is distinctive of the science and that biology is an inherently autonomous discipline. » (1961, p. 398-9)

Le premier argument rapporté (et rejeté) par cet auteur en faveur de l'indépendance de la biologie reposait justement sur l'importance des explications téléologiques dans cette science — alors qu'elles ne jouaient aucun rôle en physique ni ne pouvaient en être déduites ou dérivées —, et sur le fait que les phénomènes expliqués avaient eux-mêmes *prima facie* un caractère télique ou finaliste (*purposive character*), puisque les organismes sont capables d'auto-régulation, d'auto-maintient et d'auto-reproduction, et leurs activités sont apparemment dirigées vers des buts situés dans le futur — phénomènes n'ayant pas d'équivalent en physique

ni ne pouvant être expliqués par elle, selon cet argument. On comprend dès lors l'importance que les auteurs accordaient à l'analyse des explications téléologiques et à leur interprétation, traduction ou reformulation en d'autres termes, et à la définition de notions comme celle de fonction.

Le deuxième argument reposait sur l'idée qu'un organisme est un tout intégré dont les parties influent les unes sur les autres et dont le comportement régule et est à la fois régulé par l'activité de l'organisme conçu comme un tout, de sorte qu'il ne peut pas être analysé comme n'importe quel autre système physique composé de parties indépendantes. On peut faire remonter cette idée au moins jusqu'à Kant dans la *Critique de la faculté de juger*, lorsqu'il disait que « *dans un tel produit de la nature toute partie, tout de même qu'elle n'existe que par toutes les autres, est aussi conçue comme existant pour les autres parties et pour le tout* » (1790, paragr. 65). À l'époque de Nagel, cette idée s'inscrivait dans le cadre du débat engagé entre le mécanisme — ou réductionnisme, représenté notamment par la biologie moléculaire, — et l'organicisme — ou émergentisme, représenté par la biologie théorique — ; un débat qui s'est dissipé dans la seconde moitié du XX^e^ siècle mais qui, d'une certaine manière, est toujours actif (Etxeberria, 2006; Nicholson, 2014; Saborido, 2012).

On pourrait ajouter un troisième argument qui tient au caractère historique des phénomènes et des explications biologiques. Au début du XIX^e^ siècle, analyse Jean Gayon (1993, p. 33), lorsque apparaît le terme « biologie », deux terrains empirico-conceptuels sont assignés à cette nouvelle science : la transformation des espèces et la physico-chimie de la vie. Deux siècles plus tard, la situation de la biologie est en grande partie inchangée :

> Remarquons avec le recul du temps que ces deux engagements théoriques ont conduit à deux visions opposées de la place des sciences de la vie dans l'architectonique contemporaine des sciences de la nature. La physico-chimie de la vie va manifestement dans le sens d'une interprétation de ces sciences comme étant des provinces ou annexes des sciences générales de la matière. La théorie de l'évolution, tout particulièrement dans sa version darwinienne, avec la place immense que celle-ci réserve à la causalité et la contingence historiques, va au contraire dans le sens d'une autonomie des sciences biologiques. (Gayon, 1993, p. 33-34)

Ces deux visions se manifestent clairement dans le débat sur les fonctions biologiques tel qu'il se développe à partir des années '70 à travers les deux approches qui vont finir par se consolider et perdurer jusqu'à nos jours. D'un côté, une approche systémique qui conçoit les objets de la biologie comme des systèmes complexes parmi d'autres et dont les outils d'analyse ne distinguent pas le vivant de l'inerte. De l'autre, une approche étiologique qui se concentre de plus en plus sur la

dimension historique des phénomènes biologiques et sur la théorie darwinienne de l'évolution.

L'autonomie de la biologie a continué d'être discutée au cours des années suivantes (Ayala, 1972; Kitcher, 1984; Mayr, 1996; Rosenberg, 1985; Ruse, 1973b), mais ce n'est plus une question centrale du débat sur la téléologie et les fonctions. Toutefois, elle n'a pas disparu de la discussion (Lange, 2004; Mayr, 2007; Rosenberg, 2001a, 2001b) et certains sont en train de renouveler la question depuis une perspective différente :

> « In discussions over the last decades the special status of biology among the sciences was attributed to the central place of evolutionary theory in biology rather than to teleology. But, there can be, and for centuries there has been, if not a 'biology' then at least biological thinking without evolutionary theory. Therefore, it does not hold true that nothing in biology makes sense except in the light of evolution. In contrast to this, there never has been and never will be biology without teleological dimensions. This is because teleology, in a certain sense, is deeply rooted in the descriptive language of biology. Most biological objects do not even exist as definite entities apart from the teleological perspective. » (Toepfer, 2012, p. 118)

Depuis la seconde moitié des années '70, une partie importante de la littérature peut être conçue comme un travail interne de sophistication des approches en présence. Une autre partie se compose de formulations visant à les unifier ou à proposer des conceptions alternatives. De façon générale, l'un des objectifs de la discussion demeure celui de comprendre l'activité et le discours des biologistes quand ils recourent au concept de fonction et aux explications téléologiques.

Parmi les objectifs particuliers, l'un des plus importants est sans doute celui de la naturalisation de l'esprit ou, plus précisément, de son contenu. Il s'agissait de montrer que les capacités représentationnelles des états mentaux étaient compatibles avec le monde tel que nous le donnent à connaître les sciences de la nature. Le programme « téléosémantique », apparu dans les années '80, visait à résoudre ce problème en s'appuyant sur la notion téléologique de fonction entendue en termes de sélection naturelle (Dretske, 1986, 1988; Millikan, 1984; Papineau, 1984, 1987). En effet, selon les théories téléologiques du contenu mental, ce qu'une représentation représente dépend des fonctions des systèmes qui produisent ou emploient cette représentation, et la notion de fonction pertinente ici est celle que la biologie attribue aux traits des organismes (Neander, 2012b), d'où l'importance d'analyser correctement la notion de fonction, car elle devait faire le lien entre les phénomènes sémantiques et les phénomènes naturels.

Depuis les années 2000, la téléosémantique semble elle-aussi avoir perdu de son élan (Godfrey-Smith, 2004; Longy, 2015; Perlman, 2002) mais elle n'a pas disparu complètement du débat sur les fonctions biologiques (M. Abrams, 2005; Häggqvist, 2013; Macdonald & Papineau, 2006; Nanay, 2014; Neander, 2007).

Le programme téléosémantique visait une naturalisation en deux temps : d'abord, réduire ou analyser le contenu mental en termes de fonctions biologiques ; ensuite, réduire ou analyser ces dernières en termes non téléologiques. En ce qui concerne la première étape, le commun dénominateur aux théories téléologiques du contenu mental, d'après Neander (2012b), est l'idée que les normes psycho-sémantiques sont dérivables en dernière instance à partir des normes fonctionnelles. Cela implique que le discours fonctionnel soit effectivement normatif. Intuitivement, il semble que cela soit le cas en biologie et en médecine. Pourtant, la dimension normative des fonctions biologiques est justement l'un des points du débat depuis les années '90 et l'une des pierres d'achoppement de leur naturalisation (Amundson, 2000; Bedau, 1991; Boorse, 2002; W. Christensen, 2012; P. S. Davies, 2001; Ferguson, 2007; Fitzpatrick, 2000; Hardcastle, 2002; Krohs, 2010; McLaughlin, 2001, 2009; Molina Pérez, 2006; Mossio, Saborido, & Moreno, 2010; C. Price, 1995; Wachbroit, 1994; Wouters, 2005).

D'autres questions que la téléosémantique a contribué à renouveler sont peu présentes dans les débats. On peut par exemple s'interroger sur le projet de naturalisation lui-même. Si la biologie est déjà elle-même une science de la nature, pourquoi vouloir « naturaliser » la téléologie biologique et que doit-on entendre par là ? Ces questions nous renvoient aux problèmes originaux de l'unité de la science, de l'autonomie de la biologie vis-à-vis de la physicochimie, et de la légitimité scientifique des concepts et des méthodes qu'elle emploie, surtout celle des explications téléofonctionnelles.

Une partie de la réponse se trouve dans la suspicion que beaucoup de philosophes et de scientifiques ressentent à l'égard de la téléologie. Larry Wright a consacré le premier chapitre de son ouvrage *Teleological Explanations* (1976) à recenser les critiques formulées à son encontre. La première est l'identification des explications téléologiques avec des formes d'explication caduques issues de croyances n'ayant pas leur place dans les sciences, comme le vitalisme, l'animisme et le créationnisme :

> « The antipathy toward teleological explanations and conceptualizations, which is felt by many philosophers, scientists, and historians, stems more or less directly from the horror stories surrounding the deployment of these concepts in certain classic contexts. Arguments for vitalism, panpsychism, and a divine creator have sometimes involved teleological conceptualizations in dubious or openly fraudulent methodological practice and associated them with discredited scientific theories. » (L. Wright, 1976, p. 7)

À partir d'exemples tirés de textes d'histoire, de botanique et de psychologie, il dégage trois autres critiques, à savoir que les explications téléologiques impliquent une causalité rétrograde, qu'elles attribuent des états mentaux à des choses qui n'en ont pas et qu'elles constituent un obstacle pour la recherche scientifique :

> « [Three allegations concerning the nature of teleological explanations have been made.] These are, (a) that teleological explanations reverse the orthodox order of cause and effect, (b) that teleological explanations involve the illicit attribution of human mental characteristics to things other than human beings, and (c) that accepting teleological explanations would bring scientific research to a halt, at least in some fields. » (L. Wright, 1976, p. 10)

Ces critiques n'ont pas disparu. Encore aujourd'hui, des biologistes continuent de penser que les descriptions et les explications téléologiques sont erronées, préscientifiques et préjudiciables. Par exemple, Paul Kramer (1998) associe la téléologie à l'intentionnalité et dénonce l'usage de métaphores téléologiques en biologie : elles sont, dit-il, l'expression d'un manque de rigueur dans la pensée et dans l'écriture, une façon de parler qu'il convient d'éviter, car elle prête à confusions. David Hanke (2004) va plus loin en affirmant que la finalité (*purpose*) est entièrement subjective et n'a pas d'existence réelle en dehors de l'esprit qui la pense. D'après lui, attribuer une fin ou une fonction à un objet biologique serait une illusion résultant du fait que les activités humaines sont largement intentionnelles (*purposive*) et que nous raisonnons à propos des êtres vivants comme si c'étaient des artefacts, ce qui a des conséquences néfastes pour la science y compris lorsque la téléologie biologique est tenue pour une simple métaphore :

> « Teleology, the biologist's crutch, is bad no so much because it's lazy and wrong (which it is) but because it is straitjacket for the mind, restricting truly creative scientific thinking. Attributing function tends to foreclose further consideration of the involvement of the gene, the protein, the membrane, the cell, or the organ, in other processes. It encourages us to split any entity into sealed compartments with different "functions" instead of seeing the connections, instead of realizing it as an integrated whole. It's time to stop expecting that any part of the living world can be defined by function, however seductive or merely comfortable that feels. Feelings are deceiving. » (Hanke, 2004, p. 155)

À grands traits, on peut résumer les critiques de la manière suivante. Pour certains, la téléologie biologique implique une intentionnalité (naturelle ou divine, littérale ou métaphorique) ou une prédétermination de l'avenir. Pour d'autres, elle implique une forme invalide de la causalité (surnaturelle, rétrograde, descendante, holistique) ou bien repose sur des croyances métaphysiques et des théories obsolètes (vitalisme). Ces

critiques admettent deux interprétations : ontologique et méthodologique. La première consiste à dire que les fonctions et les fins, entendues dans un sens ou dans l'autre, n'existent pas dans la nature. La seconde est que les attributions et les explications téléofonctionnelles ne sont pas acceptables dans le cadre des sciences de la nature.

Face à ces critiques, les partisans de la naturalisation des fonctions se demandent, à l'instar de John Bigelow et de Robert Pargetter (1987, p. 182), « quel rôle les fonctions peuvent avoir dans une description purement scientifique du monde, et comment elles peuvent être "placées" dans le cadre de la science actuelle ». Que doit-on entendre par là ? Dans le contexte du programme téléosémantique, cela veut dire, suivant Neander (2012b), que les fonctions sont acceptables en biologie à condition que leur analyse soit « consistante avec l'affirmation selon laquelle le mobilier fondamental de l'univers n'est rien d'autre que ce que les sciences de la nature décrivent », à savoir l'ontologie des sciences physico-chimiques. En termes épistémologiques, cela veut dire, suivant Papineau (2005), que « les explications téléologiques sont en réalité des explications causales déguisées ».

Dans la taxonomie élaborée par Mark Perlman (2004), presque toutes les contributions au débat sur les fonctions depuis les années '70 sont classées dans la catégorie des théories naturalistes et réductionnistes — au sens où elles analysent et font reposer la téléologie et les fonctions sur des propriétés naturelles plus fondamentales. Au-delà des détails de sa classification et de la notion de réduction qu'il emploie, avec laquelle on peut ne pas être d'accord, ce que nous voulons souligner c'est le fait que la majorité des contributions au débat partagent une même perspective générale et que l'essentiel de la discussion entre elles porte sur la manière soit de réduire, soit de traduire, soit d'expliquer les fonctions biologiques en termes de mécanismes ou de relations causales.

C'est cet aspect qui nous intéresse au premier abord, c'est-à-dire la façon dont les différentes conceptions tentent de naturaliser les fonctions. Nous allons donc les examiner depuis cette perspective. Cela étant, au-delà des détails particuliers, nous voulons comprendre pourquoi le débat lui-même persiste, c'est-à-dire pourquoi ces conceptions, après quarante ans de discussions, demeurent problématiques.

On peut évidemment penser qu'il n'y a là rien d'anormal. D'abord parce que le débat est relativement récent si on le compare avec d'autres débats philosophiques, par exemple celui concernant la notion de « justice ». Ensuite parce que les progrès théoriques sont souvent très lents, aussi bien en philosophie que dans d'autres disciplines. Toutefois, on peut aussi penser que le débat n'est pas en train d'évoluer ni de progresser vers une solution quelconque, mais plutôt en train de se figer dans un antagonisme stérile (Lewens, 2000), car en dépit de la sophistication croissante des approches concurrentes, certaines des critiques de fond qu'elles se sont mutuellement lancées n'ont toujours pas reçu de

réponse satisfaisante. On peut même se demander, avec Perlman (2009, p. 18), si la théorie des fonctions se trouve aujourd'hui dans ce que Kuhn appelait la phase chaotique pré-paradigmatique des sciences, ou s'il y a une opposition entre deux paradigmes concurrents, ou encore si le paradigme en vigueur n'est autre que le naturalisme, de sorte que la prolifération des conceptions naturalistes serait seulement l'expression philosophique de la « science normale ».

Nous préférons quant à nous envisager la nature et l'issue du débat à la lumière des quatre hypothèses suivantes :

1. Polémique : Les différentes approches des fonctions biologiques sont incompatibles et rivales, et le débat doit se solder par la victoire de l'une sur les autres, de même que l'héliocentrisme de Copernic s'est imposé finalement au géocentrisme de Ptolémée.

2. Pluralisme : Les différentes approches ne sont pas concurrentes ni mutuellement réductibles, car elles ne répondent pas aux mêmes questions ni ne correspondent aux mêmes objets et méthodes ; le débat doit donc finir par admettre la pluralité des concepts de fonction.

3. Unification : Les différentes approches sont insatisfaisantes, car elles sont incomplètes et limitées, mais elles peuvent être rassemblées dans une approche unique plus générale, de même que les théories ondulatoire et corpusculaire de la lumière.

4. Dépassement : Les différentes approches sont insatisfaisantes, car elles partent de prémisses incorrectes, à savoir par exemple que les explications téléologiques sont des explications causales déguisées et que les fonctions biologiques doivent être analysées en termes de mécanismes et de relations causales.

L'avenir nous dira laquelle de ces hypothèses est correcte, mais nous essaierons ici de montrer que la première est moins plausible que les trois autres et que ces dernières sont liées. Nous verrons qu'il y a effectivement une rivalité entre différentes conceptions (voir par exemple Ariew, Cummins, & Perlman, 2002) qui se traduit par un échange d'objections et de contre-exemples visant la ligne de flottaison de l'« adversaire », mais il y a de bonnes raisons de penser qu'elles répondent en réalité à des besoins explicatifs différents, de sorte qu'il n'y aurait pas de raison de les opposer (Amundson & Lauder, 1994; Caponi, 2001b; Godfrey-Smith, 1993; Millikan, 2002; Preston, 1998; Sober, 1993). La diversité des approches et des concepts fonctionnels constitue une richesse — essentielle pour formuler de nouvelles hypothèses et généralisations — dont la biologie ne devrait pas chercher à se priver (Brandon, 2013).

De plus, tout en admettant la pluralité des besoins, nous essaierons de montrer que les différentes conceptions sont finalement assez complémentaires et qu'il n'y a pas lieu de les tenir pour incompatibles. Au

contraire, l'unification des approches pourrait aider à résoudre certaines des difficultés dont elles souffrent individuellement. Plusieurs tentatives d'unification ont déjà été formulées dans les années '90 (par exemple Buller, 1998; Griffiths, 1993; Kitcher, 1993; Walsh & Ariew, 1996). Plus récemment, l'approche dite organisationnelle apparaît comme une alternative prometteuse (W. D. Christensen, 1996; McLaughlin, 2001; Mossio, Saborido, & Moreno, 2009; Schlosser, 1998; Toepfer, 2012). Ces tentatives d'unification portent seulement sur les fonctions biologiques. Nous allons quant à nous proposer une caractérisation très générale du concept de fonction qui n'est pas limitée à la biologie, mais qui est compatible avec des formulations spécifiques à un domaine.

Les trois premières hypothèses soulèvent une série de questions ayant trait à la nature et à l'unicité du concept de fonction. Quand les biologistes, les médecins, les psychologues, les anthropologues, les archéologues, les ingénieurs et les profanes parlent de fonctions, parlent-ils de la même chose ou de choses différentes ? Quand un biologiste évolutionniste et un physiologiste actuels attribuent au cœur la fonction de pomper le sang, emploient-ils le même concept ou des concepts homonymes ? Est-ce le même que celui employé par William Harvey au XVII^e^ s. ? Ces questions qui demeurent généralement implicites dans la discussion pourraient être à l'origine d'une confusion. En effet, s'il s'avérait que tous les auteurs ne s'intéressent pas au(x) même(s) concept(s), leurs analyses ne seraient pas comparables. Il est donc possible que certains désaccords ayant alimenté le débat au cours des quatre dernières décennies soient en réalité le produit d'un malentendu.

En ce qui concerne la quatrième hypothèse, nous allons essayer de montrer tout d'abord que les principales difficultés que rencontrent les tentatives de naturalisation des fonctions biologiques reposent sur des présupposés ontologiques et méthodologiques communs aux différentes approches. Nous montrerons ensuite qu'il est possible de dépasser ces présupposés et de proposer une caractérisation générale du concept de fonction qui soit téléologique, mais qui ne soit aucunement liée à l'intentionnalité ni à des formes invalides de la causalité. De plus, nous verrons que les raisonnements téléologique et causal sont à la fois différents, autonomes et complémentaires, et que les deux sont également importants pour la biologie. Par conséquent, l'analyse des fonctions en termes de causes nous semble à la fois non nécessaire, inutile et préjudiciable.

Nous n'allons pas aborder le problème de la téléosémantique ni nous pencher en détail sur des questions spécifiques à la biologie évolutive pouvant affecter la formulation de la théorie étiologique, comme les discussions récentes sur les mécanismes de sélection, variation, etc. (voir Huneman, 2013b). Nous n'allons pas non plus nous pencher sur d'autres disciplines de la biologie, comme la génétique ou l'écologie, bien qu'elles posent des problèmes spécifiques intéressants (Bouchard, 2013; Jax, 2005; Nunes-Neto, Moreno, & El-Hani, 2014) et bien que la seconde ait

été récemment au centre d'une controverse autour de l'ADN dit « poubelle » et de la définition du concept de fonction (Doolittle, 2013; Doolittle, Brunet, Linquist, & Gregory, 2014; Elliott, Linquist, & Gregory, 2014; Germain, Ratti, & Boem, 2014; Graur et al., 2013; Kellis et al., 2014; Moran, 2014).

En revanche, nous allons nous demander ce qu'ont de particulier les objets de la biologie pour être porteurs de fonctions tandis que d'autres objets naturels ne le sont pas. Autrement dit, pourquoi avons-nous tendance à attribuer des fonctions à certains objets naturels et pas à d'autres ? Cette double question soulève le problème de la distinction entre les uns et les autres : est-elle dans les objets eux-mêmes, c'est-à-dire dans la nature, ou plutôt dans le regard, c'est-à-dire dans les méthodes scientifiques ou dans nos structures cognitives ?

De plus, si les fonctions sont des propriétés objectives des traits biologiques, on peut se demander comment et à quel moment elles furent acquises au cours de la transition de l'inerte au vivant. En effet, aujourd'hui, grâce aux progrès de la biologie synthétique, cette question théorique acquiert soudain une dimension empirique et expérimentale. Nous allons donc examiner les rapports entre la vie ou le vivant et la fonctionnalité, et mettre en parallèle les deux débats.

Par ailleurs, on peut se demander quels rapports entretiennent la vie, les fonctions et les valeurs. D'un côté, les êtres vivants se distinguent des autres objets naturels parce que nous leur attribuons des concepts normatifs et valoratifs (santé, maladie, mort, adaptation, *fitness*, succès reproductif, etc.). De l'autre, il semble que nous n'attribuions jamais de fonctions aux effets ni aux capacités délétères des traits organiques. Dans quelle mesure cette dimension axiologique est-elle essentielle à la biologie (voir Canguilhem, 1966) ? Dans quelle mesure est-elle essentielle à la téléologie (voir Bedau, 1992b; McLaughlin, 2001; Searle, 1995; Woodfield, 1976) ? Et dans quelle mesure peut-elle être naturalisée (voir Bedau, 1991; Boorse, 1977; Moreno & Mossio, 2015) ?

La tendance à attribuer des fonctions à certains objets et pas à d'autres, ainsi que la tendance à expliquer certains phénomènes en termes téléologiques, peuvent être abordées depuis la perspective psychologique. Il existe en effet d'intenses recherches en psychologie cognitive et du développement, depuis les années '90 et 2000, autour de la biologie naïve (*folk biology*) et de la pensée téléofonctionnelle, aussi bien chez l'enfant que chez l'adulte. Il nous semble que les résultats empiriques obtenus dans ce domaine pourraient fertiliser le débat philosophique. Hélas, il n'y a presque pas de contact entre les deux. Nous allons essayer de combler cette lacune en examinant les études les plus pertinentes pour nous.

Finalement, étant donné que les artefacts sont aussi des objets porteurs de fonctions, et puisque nous cherchons une caractérisation aussi générale que possible de ce concept, nous allons nous intéresser aussi au rapport entre les fonctions biologiques et techniques. Depuis la

fin des années '90, les rapports entre les organismes et les artefacts sont justement au centre de nombreuses discussions pertinentes pour le débat sur les fonctions (voir Gayon & de Ricqlès, 2010; Huneman, 2013a; Krohs & Kroes, 2009; Lewens, 2004).

Objectifs poursuivis

Nous voulons savoir tout d'abord si derrière l'apparente rivalité des différentes conceptions des fonctions se cache un pluralisme irréductible ou une unité sous-jacente. L'hypothèse est qu'il existe en biologie des besoins explicatifs différents auxquels peuvent répondre des concepts de fonction différents, mais avec une base commune. Pensons pour un instant au moteur à explosion : les camions de transport et les voitures de course ont des moteurs différents qui répondent à des besoins spécifiques, mais ils reposent sur les mêmes principes généraux que l'on peut identifier. De manière analogue, nous voulons identifier les principes généraux et les intuitions fondamentales communes aux différentes conceptions impliquées dans le débat. Nous essaierons ensuite de trouver une formulation simple, précise et générale qui montre l'unité derrière la diversité et qui permette de concilier les principales approches.

Contrairement aux tentatives d'unification précédentes, nous ne prétendons pas que notre formulation soit directement applicable par les biologistes. En premier lieu, parce qu'elle est peut-être trop abstraite. En second lieu, parce que les problèmes conceptuels et leurs solutions sont intéressants aux yeux des philosophes, mais plus rarement à ceux des scientifiques. En troisième lieu, parce qu'il n'y a pas de solution miracle pour ce problème, c'est-à-dire pas d'analyse ni de définition unique qui rende compte de tous les cas possibles, qui évite tous les contre-exemples imaginables et qui arrive à concilier toutes les postures. (De plus, comme dit Millikan (2002, p. 122), il faut arrêter de croire que les notions de fonction ont des frontières bien précises et qu'il existe une façon correcte de les employer attendant d'être découverte.) En quatrième lieu, parce que notre but n'est pas de réformer ni de légiférer, mais de poser un jalon pour notre investigation sur la légitimité des explications téléologiques et fonctionnelles.

L'une des manifestations de l'unité des fonctions est la continuité et la stabilité remarquable de ce concept depuis Aristote jusqu'à Millikan, en passant par Harvey, Darwin, Paley et Kant (voir Ariew, 2002; Cameron, 2000; Lennox, 2010). Nous pensons personnellement qu'elle va au-delà de la biologie et s'étend aux sciences humaines et aux artefacts. Nous ne sommes pas seuls sur ce point. Plusieurs auteurs, depuis des perspectives théoriques et méthodologiques différentes ont proposé des définitions suffisamment larges et abstraites pour couvrir les fonctions biologiques et techniques (par exemple Kitcher, 1993; Millikan, 1984;

Preston, 1998; Wimsatt, 1972; Woodfield, 1976; L. Wright, 1973). D'autres, comme Bigelow & Pargetter (1987, p. 194), expriment le sentiment d'une unité sous-jacente à toutes les fonctions, indépendamment de la nature biologique ou technique des traits, mais leur définition ne rend pas compte de cette idée. Nanay (2010) propose de repenser les fonctions techniques à la lumière de sa théorie des fonctions biologiques. Pour sa part, Neander (1991a, p. 175) affirme que sa théorie pourrait être formulée en termes de sélection « tout court », de manière qu'elle soit compatible avec les fonctions des artefacts et être neutre vis-à-vis de la polémique darwinisme/créationnisme — ce qui représenterait selon elle un double avantage et correspondrait à la notion de « fonction propre » du langage ordinaire dont la notion biologique est dérivée, — mais que les particularités de la sélection naturelle imposent une analyse plus spécifique. Toutefois, l'examen de la littérature semble indiquer que notre posture est minoritaire, car la plupart des auteurs limitent leurs analyses au domaine biologique et quelques-uns expriment leurs objections contre une telle unification (voir par exemple Saborido, 2012, § 1.2.2).

Mark Perlman (2009) recommandait à ceux qui se lanceraient dans le développement d'une théorie des téléofonctions qu'ils évitent de tracer une ligne inflexible entre les fonctions naturelles et celles des artefacts, car il n'y a pas de différence de principe entre les deux. Monod avançait un argument similaire dans l'introduction de *Le Hasard et la nécessité* (1970). L'idée générale défendue par ces auteurs est que les espèces animales et végétales domestiquées sont des objets naturels et artificiels à la fois. De même, dit Risto Hilpinen (1993), les bactéries et autres microorganismes qui ont été intentionnellement manufacturés pour atteindre un objectif donné, comme dégrader le pétrole, possèdent les caractéristiques principales des artefacts, bien qu'ils appartiennent par ailleurs à un genre naturel (celui des bactéries). De façon générale, il n'y a pas de ligne de démarcation bien tranchée entre ce qui appartient à la nature et ce qui relève de l'art ou de la technique, et la distinction entre les deux varie selon les époques et les cultures (Bensaude-Vincent & Newman, 2007; Ehrman & Grosseield, 1980).

Dan Sperber (2007) appelle les espèces domestiquées des artefacts biologiques culturels qui gomment la dichotomie entre les fonctions entendues comme effets sélectionnés (*selected effects*) et comme effets voulus (*intended effects*). En effet, les graines de plantes cultivées comme le blé ont à la fois une fonction biologique (la reproduction de la plante) et une fonction en tant qu'artefact (l'alimentation humaine) laquelle est à la fois une fonction culturelle. Or, de nombreuses plantes à graines ont évolué de manière à favoriser leur propagation grâce aux animaux non-humains. Le fait qu'elles aient été exploitées culturellement par les humains pour leur alimentation, avec la révolution agricole, a contribué au succès reproductif des espèces et des variétés qui ont pu évoluer dans ce sens (toutes les espèces n'étant pas domesticables ou pas dans la même

mesure). Par conséquent, dit Sperber, on peut considérer l'alimentation humaine et la reproduction comme étant des fonctions culturelles et biologiques à la fois :

> « feeding humans became a biological teleofunction of the seeds, that is, an effect that contributed to the greater reproductive success of varieties of cereal providing better food. Both the feeding and the reproduction function of seeds are simultaneously biological and cultural/artifactual functions of cultivated cereal. The plants take biological advantage of their cultural functions and humans exploit culturally, and more specifically economically, some of the biological functions of the plants. The has been a co-evolution of the plants and of their cultural role. Human culture has adapted to cereal biology just as cereals have adapted to human culture. » (Sperber, 2007, p. 133-134)

D'autres auteurs vont même jusqu'à dire que la domestication des plantes et des animaux est un processus bidirectionnel de co-évolution, de sorte que les humains ont en quelque sorte été domestiqués par les plantes (Harari, 2015; Jackson, 1996) et par les chiens (Hare & Woods, 2013; Sperber, 2007) — après que les ancêtres de ces derniers se fussent auto-domestiqués.

Il convient de remarquer, en effet, que les humains ne sont pas la seule espèce animale capable d'en domestiquer d'autres. Les fourmis, par exemple, pratiquent l'agriculture (Mueller, Rehner, & Schultz, 1998) et l'élevage (Oliver, Mashanova, Leather, Cook, & Jansen, 2007). Certes, cette domestication n'est pas intentionnelle, mais celle pratiquée par les humains ne l'est pas non plus entièrement, car comme l'avait déjà signalé Darwin, certains des traits désirables des espèces domestiquées sont des résultats involontaires des pratiques d'élevage et de culture ; ils n'ont pas été intentionnellement recherchés.

Les humains ne sont pas non plus les seuls à fabriquer des objets fonctionnels. De nombreuses autres espèces animales produisent elles-aussi des objets auxquels nous n'hésitons pas à attribuer des fonctions, comme la toile des araignées, les ruches des abeilles, les barrages des castors, les tonnelles décorées des oiseaux jardiniers, etc. Certains auteurs proposent même d'étendre la notion d'artefact à ces objets manufacturés, bien qu'ils ne soient pas forcément le produit d'une activité intentionnelle ni d'une transmission culturelle (J. L. Gould, 2007).

Les exemples précédents ne remettent pas forcément en question la distinction tracée entre les fonctions biologiques et celles des artefacts. Les toiles d'araignées et autres artefacts animaux (si l'on accepte de les considérer comme tels) pourraient en effet se voir attribuer une fonction biologique, bien qu'ils ne soient pas des traits organiques ni des comportements, mais des produits non intentionnels de ces comportements. On pourrait par exemple considérer, avec Richard Dawkins (2016), que la

notion de phénotype doit inclure la toile des araignées et d'autres objets externes dans la mesure où ce sont eux-aussi des effets des gènes. Cependant, tous les artefacts que l'on trouve dans le monde naturel ne sont pas déterminés génétiquement.

Certains animaux non-humains sont capables d'utiliser des objets naturels comme outils, et même de les modifier ou de les manufacturer, comme ces chimpanzés qui fabriquent des sortes de lances avec lesquelles ils chassent d'autres singes (Pruetz et al., 2015). Ces objets ont des propriétés fonctionnelles que les animaux peuvent identifier, sélectionner et modifier (Clair & Rutz, 2013; Hauser, 2002; Herrmann, Wobber, & Call, 2008; Ruiz & Santos, 2013). Dans la mesure où ils résultent de l'activité d'animaux sauvages en milieu naturel, on peut considérer que ces outils sont des objets naturels. Cependant, dans la mesure où ils sont manipulés ou manufacturés, ou détournés de leur fonction naturelle en vue d'une certaine fin, on peut les considérer comme des objets artificiels, et à plus forte raison lorsque le comportement en question n'est pas hérité génétiquement mais transmis culturellement.

D'après certains auteurs, un objet ne peut pas être un outil à moins qu'il ne soit employé à dessein (*purposively*), de sorte que l'utilisation d'outils implique une finalité (*purposiveness*), mais pas forcément une conscience ni une compréhension causale de la part de l'animal (Shumaker, Walkup, & Beck, 2011; St Amant & Horton, 2008). S'il en est ainsi, alors il n'est pas absurde d'attribuer aux outils animaux et à leurs parties une fonction intentionnelle. D'après Risto Hilpinen (2011), l'activité intentionnelle (*intentional agency*) ne se limite pas aux êtres humains.

Cela étant dit, l'utilisation et la production d'outils et l'évolution des caractéristiques fonctionnelles de ces derniers peut être liée soit à la sélection naturelle des comportements correspondants, soit à leur transmission culturelle, soit à plusieurs mécanismes à la fois. Quoi qu'il en soit, les outils animaux viennent brouiller la distinction entre les fonctions biologiques et celles des artefacts, surtout si l'on considère que les humains sont une espèce de primates parmi d'autres.

Après avoir examiné attentivement les frontières présumées entre les fonctions biologiques, intentionnelles et culturelles — du point de vue de la théorie étiologique —, Françoise Longy (2009, 2010) soutient elle aussi, depuis une perspective différente, que ces frontières n'existent pas et que l'on trouve partout un *continuum*, avec des fonctions mixtes et intermédiaires entre plusieurs catégories. Elle en conclut que l'on devrait abandonner l'idée d'une distinction entre celles des objets naturels et celles des artefacts :

> « I argue that at least as far as material and biological entities are concerned, we should renounce the idea of putting their functions into different ontological categories. This therefore implies that we must abandon the idea that we should or could have different accounts of functions of material entities depending either on the

> nature of the entities (natural or artificial, inert or living), or on the origin of the functions (selection or intention). » (Longy, 2009, p. 54)

Contrairement à Perlman, qui est partisan du pluralisme, Longy appelle de ses vœux une notion de téléofonction unique plus abstraite que celles fournies par les théories étiologiques actuelles ; une notion qui ignorerait la question des origines :

> « A more abstract notion of "teleofunction," which would ignore the question of origins, could help us to understand why identifying functions is so useful when faced with relatively simple high-level phenomena that depend on complex mechanisms operating at different levels. » (Longy, 2009, p. 65)

Nous partageons ces conclusions et les faisons nôtres. Nous viserons donc une formulation abstraite de la notion de fonction qui dépasse la distinction entre les objets biologiques et les artefacts. De plus, tout au long de ce travail, nous utiliserons presque indistinctement des exemples tirés des domaines biologique et technique pour discuter un point ou illustrer un argument. Cela étant, notre attention sera centrée principalement sur le débat en philosophie de la biologie, de sorte que nous n'examinerons pas en détail les problèmes spécifiques aux artefacts ni aux sciences humaines.

Par ailleurs, nous reviendrons à plusieurs reprises sur l'idée que les fonctions et le vivant semblent être intimement liés. En effet, parmi les objets et les phénomènes naturels, nous n'attribuons généralement de fonctions qu'à ceux qui relèvent de la biologie, c'est-à-dire à ceux qui appartiennent à des êtres vivants ou qui leur sont associés[1], tandis que tous les autres objets naturels sont apparemment dénués de fonctions. Or, cette idée, dont il faudra examiner la valeur, n'est pas incompatible avec la recherche d'une formulation générale qui inclue les artefacts, non seulement parce que les frontières entre le naturel et l'artificiel sont floues, comme nous venons de le dire, mais aussi parce que *tous* les objets auxquels nous attribuons des fonctions sont liés au phénomène de la vie. Les artefacts eux-mêmes sont en effet des produits de l'activité d'êtres vivants. Par conséquent, toutes les attributions fonctionnelles dans le cadre des sciences de la nature portent soit sur des parties d'êtres vivants (structures, processus, systèmes), soit sur des actions ou comportements d'êtres vivants, soit sur des produits de ces actions ou comportements.

Cela ne veut pas dire pour autant que la fonctionnalité et la vie soient liées de façon significative. Or, il nous semble important de faire la lumière sur cette question. Nous allons donc nous pencher sur les travaux

1 Nous attribuons des fonctions non seulement aux structures, processus et comportements des organismes cellulaires, mais aussi, par exemple, aux gènes des virus et des plasmides, lesquels sont généralement considérés comme non-vivants parce qu'ils n'ont pas de métabolisme propre.

qui portent sur les conditions de possibilité du vivant, que ce soit dans le domaine des cellules minimales, en biologie synthétique, ou dans celui des recherches sur les origines de la vie. Si la vie et les fonctions sont liés, ces travaux devraient nous aider à comprendre comment et pourquoi. À partir de là, nous viserons à comprendre si les fonctions biologiques et celles des artefacts le sont dans le même sens ou s'il s'agit de notions différentes. Notre hypothèse est que les fonctions ne sont pas liées à la nature des objets eux-mêmes — c'est-à-dire par exemple au fait qu'ils soient vivants ou qu'ils aient un certain type d'organisation interne —, ni à leur origine causale — c'est-à-dire par exemple à la sélection naturelle ou à un acte de création intentionnelle —, mais plutôt au regard que nous portons sur eux, c'est-à-dire au type d'explication que nous employons pour en rendre compte.

Par ailleurs, nous faisons l'hypothèse que le principal obstacle pour la naturalisation des fonctions est l'idée selon laquelle les explications téléofonctionnelles seraient des explications causales déguisées et que les naturaliser consisterait à les analyser, réduire, traduire ou expliciter en termes de relations causales scientifiquement acceptables. Cette idée que nous rejetons repose sur le raisonnement suivant.

(a) La téléologie, prise telle quelle, fait référence à des formes de causalité scientifiquement invalides (rétrogrades, surnaturelles, etc.) ou inapplicables en biologie (intentionnalité).

(b) Les explications causales, comme celles que l'on trouve en physique (classique), sont la forme canonique de l'explication dans les sciences de la nature.

(c) Le seul moyen de réhabiliter la téléologie biologique et de lui faire une place au sein des sciences de la nature est de l'interpréter conformément à (b).

L'un des objectifs principaux de ce travail est de montrer qu'il existe une alternative à (c). Pour cela, nous devons proposer une interprétation qui ne soit pas problématique comme (a), ni causale comme (b). En effet, nous ne pensons pas que les explications téléofonctionnelles soient des explications causales déguisées (valides ou invalides), ni que les explications causales soient les seules explications scientifiques valables. Ces deux dernières affirmations mériteraient une justification détaillée qui excède le cadre de notre investigation sur les fonctions. Nous nous limiterons donc à essayer de montrer que les explications téléologiques sont différentes et autonomes des explications causales et intentionnelles, qu'elles ne sont pas nécessairement problématiques aux sens indiqués plus haut, qu'elles peuvent être compatibles avec des théories modernes de l'explication scientifique comme celle de James Woodward (2003) et qu'elles ont une valeur explicative pour la biologie que n'ont pas les explications causales.

Méthode de travail

Du point de vue méthodologique, nous avons le choix entre plusieurs options, parmi lesquelles se trouvent l'analyse conceptuelle, la définition théorique et l'explication philosophique.

Dans le premier cas, il s'agirait pour nous d'analyser la notion de fonction du langage ordinaire, mais par l'intermédiaire du débat philosophique. Autrement dit, en partant de l'hypothèse que les différentes conceptions philosophiques des fonctions sont incomplètes et qu'il est possible de les unifier, nous pourrions isoler les éléments essentiels de ces conceptions pour chercher à identifier un dénominateur commun. Dans une certaine mesure, c'est ce que nous allons faire dans la première et la troisième parties de ce travail, mais pas dans le cadre d'une analyse conceptuelle — ou du moins pas dans l'intention d'élucider la signification du concept ni de déterminer ses conditions d'application.

Dans le deuxième cas, il s'agirait de comprendre ce que *sont* les fonctions, indépendamment des représentations mentales que l'on peut en avoir. Dans une certaine mesure, c'est aussi ce que nous allons faire dans la deuxième partie de ce travail en interrogeant la biologie synthétique et les théories sur l'origine de la vie. En effet, si les fonctions sont des propriétés des organes, des structures et des processus du vivant, alors la compréhension des mécanismes du vivant et de ses conditions d'apparition pourrait nous permettre de mieux comprendre ce que sont ses fonctions. Pourtant, nous ne proposerons pas une définition théorique.

Les trois méthodes ont des avantages et des inconvénients. Après en avoir signalé quelques-uns, dans les paragraphes qui suivent, nous opterons finalement pour l'explication philosophique.

Analyse conceptuelle

L'un des problèmes de l'analyse conceptuelle est que derrière les désaccords qui se manifestent de façon explicite dans la discussion à propos de la signification ou des critères d'application de la notion de fonction s'en cachent d'autres, implicites, qui portent sur l'objectif même du débat et sur le résultat qui en est attendu. Nous allons voir à travers les quelques exemples suivants que tous les auteurs n'ont pas le même projet ni n'emploient les mêmes méthodes. Par conséquent, il est difficile de mesurer la portée des objections qu'ils se lancent les uns aux autres et dans quelle mesure et en quel sens une théorie ou une définition est prétendument meilleure qu'une autre (Lewens, 2000, p. 95).

On peut en principe considérer que la plupart des contributions au débat sont des analyses conceptuelles conformes à la Théorie classique des concepts entendue de la manière suivante :

> « According to the classical theory, a lexical concept *C* has definitional structure in that it is composed of simpler concepts that express necessary and sufficient conditions for falling under *C*. […] Paradigmatic conceptual analyses offer definitions of concepts that are to be tested against potential counterexamples that are identified via thought experiments. Conceptual analysis is supposed to be a distinctively a priori activity that many take to be the essence of philosophy. » (Margolis & Laurence, 2014)

En effet, la plupart des auteurs proposent une définition formulée sous forme de conditions nécessaires et suffisantes, et la plupart des critiques à l'encontre de ces définitions reposent sur des contre-exemples dictés par l'intuition.

Ainsi, Larry Wright (1973, p. 162) considère que son analyse est sur la bonne voie parce qu'elle rend compte de tous les cas envisagés par lui, tout en évitant un certain nombre de contre-exemples gênants. De plus, il élabore son analyse *a priori* et s'appuie entièrement sur l'intuition pour établir ou rendre compte des distinctions pertinentes et pour élucider la signification du concept de fonction. Andrew Woodfield (1976) est encore plus explicite dans le chapitre qu'il consacre à sa méthode. Il ne s'agit pas, dit-il, d'expliquer ce qu'est la téléologie ni de la réduire à des propriétés physiques, comme on pourrait le faire pour la notion de chaleur à partir d'une étude scientifique. Il s'agit plutôt de la définir, c'est-à-dire de rendre explicites les conditions nécessaires et suffisantes d'application de plusieurs types de phrases téléologiques du langage ordinaire. Il s'appuie pour cela sur l'intuition, et les conditions d'application sont testées en essayant d'imaginer des contre-exemples. Ce travail débouche sur un ensemble de quatre définitions — ou paraphrases — qui révèlent le sens des différents types de phrases téléologiques.

Karen Neander (1991a) revendique elle aussi l'analyse conceptuelle, qu'elle entend comme la tentative de décrire les critères d'application (implicites ou explicites) que les membres d'une communauté linguistique ont généralement à l'esprit quand ils emploient le terme analysé, mais elle rejette deux des caractéristiques centrales de la Théorie classique du concept, à savoir la recherche de conditions nécessaires et suffisantes, et la recherche de la signification. De plus, elle affirme que les intuitions et les contre-exemples ne devraient pas être considérées comme les arbitres finaux de l'analyse, car celle-ci doit être également être guidée par le rôle théorique du concept. Sur ce point, elle coïncide avec d'autres auteurs (par exemple Allen & Bekoff, 1995a; Bigelow & Pargetter, 1987; Godfrey-Smith, 1994). Ce qui ne les empêche pas de se servir eux-mêmes de contre-exemples et de conséquences contre-intuitives pour attaquer les définitions d'autres auteurs.

Neander ne se pose pas la même question que Wright et Woodfield. Ces derniers s'intéressent à la signification générale du concept, tandis que Neander s'interroge sur son rôle explicatif en biologie. De plus,

contrairement à ses prédécesseurs, Neander limite son analyse à un terme technique employé par une communauté restreinte de spécialistes. Il ne s'agit donc peut-être pas du même concept, et quand bien même ce serait le cas on ne pourrait pas évaluer leurs analyses de la même façon.

En nous limitant au domaine scientifique et au rôle théorique du concept de fonction en biologie, les choses ne sont pas beaucoup plus simples. En premier lieu, tous les auteurs ne sont pas forcément d'accord pour reconnaître aux mêmes éléments le même rôle théorique. Par exemple, Bigelow & Pargetter (1987), contrairement à Neander, considèrent que l'histoire d'un trait biologique ne joue aucun rôle explicatif et que, d'un point de vue théorique, le concept de fonction doit être analysé en termes de *propension* à la sélection et pas d'*histoire* sélective.

En second lieu, l'une des objections habituellement lancées contre l'approche de Neander est que sa définition ne rend pas compte du concept de fonction tel qu'il était employé par les médecins et les biologistes avant la théorie darwinienne, c'est-à-dire notamment par Harvey, découvreur de la circulation du sang et de la fonction cardiaque. La réponse de Neander est que les concepts évoluent et que son analyse porte sur l'usage des biologistes actuels. Or, si l'analyse conceptuelle vise à comprendre ce que les locuteurs ont à l'esprit (Neander, 1991a, p. 170), cela peut signifier que le concept actuel a changé par rapport à celui de Harvey, ou alors que le nôtre et le sien sont deux concepts différents portant le même nom. Pourtant, nous arrivons à le comprendre sans difficulté apparente. Cela peut donc aussi vouloir dire que nous partageons essentiellement le même concept, celui dont les critiques réclament justement l'analyse.[2]

En troisième lieu, même en limitant l'analyse à l'usage technique de la notion de fonction par les biologistes actuels, on peut ne pas tomber d'accord à propos de son unicité. Par exemple, Gustavo Caponi (2001a), s'appuyant sur Ernst Mayr, soutient que la biologie fonctionnelle et la biologie évolutive sont deux domaines presque autonomes, tant du point de vue méthodologique que conceptuel, et qu'il existe une différence fondamentale et irréductible entre leurs notions respectives de fonction. S'il en était ainsi, on pourrait considérer que la définition de Neander capture la notion propre à la biologie évolutive, tandis que celle de Cummins capture la notion propre à la biologie fonctionnelle. Las, Amundson & Lauder (1994) soutiennent que les deux notions sont également nécessaires aux deux domaines et qu'elles sont mutuellement irréductibles. Plus près de nous, Brian Garvey (2007, p. 125) affirme que Cummins et Neander parlent de choses différentes et que leurs analyses ne sont donc pas rivales, bien que les concepts qu'ils capturent soient tous deux très importants pour l'explication en biologie et qu'ils portent le même nom. Philip Kitcher et plusieurs autres auteurs reconnaissent

2 Pour une brève histoire du concept de fonction de Aristote à Millikan en passant par Harvey, voir James Lennox (2010).

eux-aussi la diversité des pratiques et des concepts de la biologie, mais ils estiment néanmoins possible de les concilier dans un concept unique :

> « Philosophical discussions of function have tended to pit different analyses and different intuitions against one another without noting the pluralism inherent in biological practice. On the account I have offered here, there is indeed a unity in the concept of function, expressed in the connection between function and design, but the sources of design are at least twofold and their relation to the bearers of function may be more or less direct. This means, I believe, that the insights of the main competitors, Wright's aetiological approach and Cummins's account of functional analysis, can be accommodated […]. » (Kitcher, 1993, p. 278)

Finalement, l'analyse conceptuelle exclut en quelque sorte la possibilité que les fonctions n'existent pas et que les attributions de fonctions soient fausses (Lewens, 2000, p. 97). En effet, elles ne sont pas analysées en tant que propriétés des objets de la biologie, mais en tant que représentations mentales[3]. Neander reconnaît que l'on peut analyser de la même façon les concepts de « sorcière », « entéléchie » et « phlogistique », car on peut expliquer ce que les gens pensaient à leur propos (1991a, p. 172), et peu importe que les théories qui sous-tendent ces concepts soient fausses et que ces choses elles-mêmes n'existent pas, car la vérité pertinente pour l'analyste conceptuel est ce que les gens à un moment donné *croient* (1991a, p. 178).

La méthode suivie pour déterminer ce que les gens croient n'est pas pour autant celle de la psychologie, ni de la sociologie, ni de l'histoire, c'est-à-dire celle des sciences empiriques. Il nous semble personnellement qu'elle repose en grande partie sur les intuitions des philosophes. Or, la diversité des intuitions parmi les philosophes n'est pas forcément moindre que parmi les biologistes et les profanes. Par conséquent, il est possible que le débat dégénère dans ce que Bigelow et Pargetter appellent le « bruit sourd des conflits d'intuitions » (1987, p. 196).

L'idée de se tourner vers les biologistes n'est pas partagée par tout le monde. Selon Paul S. Davies (2001), la tâche du philosophe n'est pas de montrer comment les spécialistes emploient la notion de fonction, ni ce qu'ils en disent ou pensent, mais de savoir ce qu'elles *sont*, tant du point de vue ontologique (quelle est leur place dans le monde) qu'épistémologique (comment elles s'insèrent dans le cadre général des sciences de la nature). Ruth Millikan (1989b) rejette elle aussi le recours à l'analyse conceptuelle, car le véritable problème n'est pas de savoir ce que les biologistes veulent dire, mais de comprendre les phénomènes auxquels leurs concepts font référence.

3 L'analyse conceptuelle n'implique pas nécessairement que les concepts soient des représentations mentales, mais c'est ainsi que l'entendent, nous semble-t-il, plusieurs des auteurs qui s'expriment à ce sujet.

En effet, puisque les fonctions sont un concept explicatif de la biologie, qui est une science de la nature, on peut raisonnablement penser qu'elles font référence à des propriétés ou des phénomènes naturels, c'est-à-dire à quelque chose qui existe d'une certaine manière dans le monde. Du point de vue ontologique, les tentatives de naturalisation consistent justement à expliquer de quelle manière elles existent dans le monde, comment elles dérivent d'une ontologie plus fondamentale. Dès lors, au lieu de s'intéresser à ce que disent et pensent à leur propos les gens ordinaires et les spécialistes, il faudrait se demander ce qu'elles sont.

Définition théorique

Pour répondre à cette question, Ruth Millikan rejette explicitement l'analyse conceptuelle et propose une définition théorique, comme celles que formulent les scientifiques quand ils disent que l'eau est H_2O ou que l'or est un élément chimique dont le nombre atomique est 79. Dans le cas des fonctions, il existe dit-elle un phénomène commun sous-jacent à toutes les situations dans lesquelles nous attribuons des fonctions ou des fins (*purposes*) aux choses. Ce phénomène, qu'elle appelle « fonction propre », est responsable des signes ou des propriétés qui nous poussent à penser que quelque chose a une fonction ou une fin, de même que la composition chimique de l'eau est responsable des propriétés manifestes (fluidité, transparence, saveur, etc.) qui font que nous désignons un corps comme étant de l'eau. Sa définition de « fonction propre » est donc conçue comme une définition théorique de « fonction » ou de « finalité » :

> « The definition of "proper function" is intended as a theoretical definition of function or purpose. It is an attempt to describe a unitary phenomenon that lies behind all the various sorts of cases in which we ascribe purposes or functions to things, which phenomenon normally *accounts for* the existence of the various analogies upon which applications of the notion "purpose" or "function" customarily rest. My claim is that actual body organs and systems, actual actions and purposive behaviors, artifacts, words and grammatical forms, and many customs, etc., all have proper functions, and that these proper functions correspond to their functions or purposes ordinarily so called. Further, it is *because* each of these has a proper function or set of proper functions that it has whatever marks we tend to go by in *claiming* that it has functions, a purpose, or purposes. » (Millikan, 1989b, p. 293)

Il y a plusieurs différences entre la définition théorique de l'or et celle des fonctions, outre le fait que ces dernières ne sont pas un genre naturel. L'or, l'eau, le magnétisme, l'albédo, etc., sont soit des choses tangibles, soit des phénomènes ou des propriétés dont nous n'avons pas toujours connu la nature ni l'explication, mais dont la réalité n'était pas

en question. Prenons un autre exemple. L'existence de la matière noire cosmologique est aujourd'hui discutée. Ceux qui postulent son existence, à savoir la plupart des astrophysiciens, le font parce qu'elle permettrait d'expliquer certains phénomènes que l'on ne sait pas expliquer autrement, comme la vitesse de rotation des galaxies. Ceux qui la nient pensent que ce sont les théories qu'il faut changer. Quoi qu'il en soit, la confirmation de son existence dépend non seulement de son utilité théorique, mais aussi de son éventuelle observation empirique. S'il s'agissait par exemple de particules exotiques, alors il faudrait les mettre en évidence expérimentalement, comme cela a été fait récemment avec le boson de Higgs, les ondes gravitationnelles, etc.

En ce qui concerne les téléofonctions, c'est non seulement leur existence qui est contestée, mais aussi celle des phénomènes normatifs qu'elles permettraient d'expliquer. Selon Davies (2001) et d'autres, il n'y a pas de normes dans la nature, contrairement à ce que les théories étiologiques voudraient nous faire croire, et l'apparence de leur existence est seulement d'ordre psychologique. Millikan justifie en dernière instance sa définition par son utilité théorique. Est-ce suffisant ? Probablement pas. Après tout, le phlogistique pouvait lui-aussi être justifié par son utilité théorique. De plus, certains lui reprochent que sa définition soit utile pour sa théorie de l'esprit et du langage, mais pas pour la biologie, et qu'elle s'appuie sur des intuitions plutôt que sur ce qui se passe vraiment dans cette discipline (Wouters, 2005).

La théorie étiologique défend une conception réaliste des fonctions, que ce soit par le biais d'une définition théorique ou d'une analyse conceptuelle du discours des biologistes, car attribuer une fonction, c'est faire une affirmation quand à l'histoire causale du porteur de fonction (Huneman, 2013c; Longy, 2009). La théorie du rôle causal de Cummins (1975) n'est pas réaliste, car les fonctions y sont relatives aux intérêts épistémiques de l'observateur. D'autres conceptions systémiques sont explicitement réalistes, comme celle de Davies (2001), ou peuvent être considérées comme telles (Mossio et al., 2009). De fait, hormis celle de Cummins, il nous semble que la plupart des conceptions naturalistes — pour ne pas dire toutes — reposent sur une forme ou une autre de réalisme. Nous entendons par là que les fonctions sont supposées exister dans la nature indépendamment de l'observateur, que ce soit sous la forme d'une histoire causale, de propriétés systémiques ou d'autre chose.

De manière récurrente dans ce travail, nous nous demanderons ce que *sont* les fonctions, mais nous n'avons pas l'ambition d'en proposer une définition théorique à la manière de Millikan. Dans la deuxième partie de ce travail, nous allons établir un parallèle entre le débat sur la définition des fonctions et celui sur la définition de la vie et du vivant. Ces concepts étant étroitement liés, nous voulons comparer les stratégies de naturalisation dans un cas et dans l'autre. Notre hypothèse est qu'elles rencontrent des problèmes similaires parce qu'elles reposent sur des

présupposés similaires, comme le réalisme. Or, que se passerait-il si la vie et les fonctions n'étaient pas dans la nature, mais dans le regard, c'est-à-dire dans le discours et les méthodes de la biologie ? Les attributions de fonctions seraient-elles nécessairement subjectives, c'est-à-dire arbitraires ? Les explications téléofonctionnelles auraient-elles une valeur explicative ? Nous consacrerons les quatrième et cinquième parties à répondre à ces questions.

Explication philosophique

Une autre alternative à l'analyse conceptuelle est celle que Peter Schwartz (2004, p. 143-5) appelle *explication philosophique* et qu'il tire de Carnap (1950) et de Quine (1960). Elle ne consiste pas, dit-il, à capturer la signification d'un concept ni à décrire ce que les gens ont à l'esprit quand ils l'emploient, mais à définir un nouveau terme qui joue le même rôle que l'ancien tout en évitant de poser les mêmes problèmes. Elle ne requiert pas non plus que les gens adoptent et suivent au pied de la lettre la définition du nouveau terme. Les biologistes, continue Schwartz, pourraient ainsi continuer à employer le mot « fonction » de manière ambigüe et même contradictoire, ce qui peut avoir ses avantages d'un point de vue heuristique ou métaphorique. Mais pour qu'une attribution fonctionnelle puisse être prise littéralement les biologistes devraient être disposés à adopter une définition précise et montrer que l'attribution en satisfait les conditions. De cette manière, puisqu'il ne s'agit pas de rendre compte du sens ni de l'usage actuels, l'explication philosophique ne se prononce pas quant au caractère téléologique ou pas du concept biologique en vigueur. D'après Schwartz, elle se contente de dire que l'on peut continuer à attribuer des fonctions sans avoir recours à des notions problématiques, car il est possible de formuler une définition qui ne le soit pas — ou qui du moins n'ait pas les mêmes problèmes.

Un autre aspect important de l'explication philosophique est qu'elle n'exclut pas le pluralisme. Plusieurs formulations incompatibles sont en effet envisageables, et c'est alors aux biologistes de choisir laquelle leur convient le mieux pour rendre compte du sens littéral des attributions fonctionnelles :

> « The availability of two (or more) incompatible accounts would similarly answer the original set of concerns. In this case, biologists would be free to choose either account to explain the literal meaning of their terms. As long as we are looking for a way to replace traditional, problematic concepts with new, useful ones, *two* options are as good (if not better) than one. If the analysis of "function" *was* conceptual analysis, the existence of two incompatible accounts would not be an acceptable conclusion to the discussion. » (Schwartz, 2004, p. 145)

Depuis cette perspective, les conceptions de type étiologique ou systémique pourraient être vues comme des explications philosophiques alternatives mises à disposition des biologistes, chacune répondant à des besoins particuliers.

Le principal problème que pose cette méthode est le risque de « balkanisation » conceptuelle. C'est-à-dire que par-delà la traditionnelle division entre biologie évolutive et fonctionnelle, chaque sous-domaine de ces grands domaines pourrait légitimement revendiquer un ou plusieurs concepts de fonction taillés à la mesure de ses propres besoins. Il en résulterait une multiplication incontrôlée de concepts homonymes et une grande confusion.

Cette confusion est néanmoins évitable en imaginant une version alternative du rasoir d'Ockham : « Les définitions ne doivent pas être multipliées par-delà de ce qui est nécessaire ». Suivant ce principe, s'il était possible de trouver une définition unique répondant à plusieurs besoins disciplinaires, elle serait préférable à deux définitions différentes. Par conséquent, sans remettre en question le pluralisme implicite de cette méthode, nous estimons que l'explication philosophique peut et doit être utilisée dans la mesure du possible pour unifier les approches et pas pour les diviser davantage.

La méthode de Schwartz est une solution pragmatique qui ne requiert pas d'engagement ontologique et qui admet le pluralisme et l'unification. Elle nous semble donc plus appropriée que l'analyse conceptuelle et que la définition théorique pour le travail que nous voulons réaliser ici et le type de questions que nous nous posons.

En ce qui concerne la forme, la plupart des auteurs proposent une définition des fonctions généralement formulée en termes de conditions nécessaires et suffisantes. Une telle formulation est-elle inévitable ? Peut-être pas. Pour nous, l'objectif n'est pas d'offrir une définition directement exploitable par les biologistes, mais plutôt d'apporter une solution au débat philosophique. Nous voulons montrer, dans la troisième partie de ce travail, qu'il est possible d'unifier les approches étiologique et systémique, que l'on peut dépasser la distinction naturel/artificiel, et qu'il n'est pas nécessaire d'analyser les fonctions en termes de causes pour les rendre scientifiquement acceptables. Pour cela, nous allons proposer une formulation précise en termes de conditions nécessaires (mais pas forcément suffisantes). On peut donc la considérer, suivant Françoise Longy (2013, p. 209), non pas comme une définition à part entière, mais plutôt comme une caractérisation.

Structure générale de l'argument

Ce travail est divisé en cinq parties. La première partie est consacrée au débat sur la naturalisation des fonctions dont nous avons déjà esquissé les grands traits. Nous examinerons successivement les approches étiologique (CHAP. I), systémique (CHAP. II) et mixtes (CHAP. III). La première regroupe les postures qui accordent aux fonctions une dimension temporelle, c'est-à-dire notamment celles qui les situent dans le cadre de la théorie darwinienne de l'évolution. La seconde regroupe celles qui situent les fonctions dans le cadre de l'activité d'un système, que ce soit en termes de rôle causal ou de contribution à un but. En ce qui concerne les troisièmes, elles consistent à retenir certaines des intuitions les plus intéressantes ou les plus convaincantes de chacune des deux approches précédentes pour essayer de les unifier sous une conception unique.

La deuxième partie est consacrée à l'examen des fonctions biologiques depuis la perspective des recherches sur le vivant. Nous nous intéresserons à la biologie synthétique et aux recherches sur les origines de la vie, c'est-à-dire aux disciplines qui étudient caractéristiques minimales du vivant, pour voir s'il est possible d'expliquer ou de caractériser, du point de vue scientifique, ce que sont les fonctions (CHAP. IV). Cela nous amènera à nous interroger sur la réalité et l'objectivité scientifique du vivant et de ses frontières, car si la distinction entre le vivant et l'inerte dépendait en quelque sorte de l'observateur, il pourrait en aller de même des fonctions biologiques (CHAP. V).

La troisième partie est consacrée à la caractérisation générale que nous proposons afin d'unifier les approches étiologique et systémique, de couvrir les fonctions biologiques et techniques, et de dépasser le réductionnisme causal. Nous commencerons par expliquer en quoi elle consiste et par discuter certaines de ses implications (CHAP. VI). Ensuite, nous proposerons une formulation précise de cette conception des fonctions et nous tâcherons de la justifier (CHAP. VII). Il s'agit en quelque sorte d'un retour aux sources cybernétiques de la discussion, quoique depuis une perspective différente, car notre caractérisation des fonctions les rend explicitement relatives à une fin du système contenant.

La quatrième partie est consacrée à la justification du recours à une conception non-causale de la finalité et à sa légitimité épistémologique. Nous verrons en premier lieu que le mode de raisonnement téléologique est vraisemblablement inné et autonome, du point de vue de la psychologie cognitive, vis-à-vis des modes de raisonnement causal et intentionnel (CHAP. VIII). Nous allons ensuite comparer les explications causales et téléologiques à partir d'un ensemble de cas très simples pour montrer en quoi elles consistent et en quoi elles diffèrent (CHAP. IX). Nous verrons aussi que la téléologie est innocente des charges dont on l'accuse (CHAP. X). Finalement, nous essaierons de montrer que les attributions et

les explications téléofonctionnelles en biologie ont un rôle scientifique important différent de celui des explications causales (CHAP. XI).

La cinquième partie est consacrée à trois approches alternatives — affines à la nôtre — à partir desquelles il est également possible de justifier l'attribution de fins à des objets naturels. Il s'agit des approches mentaliste (CHAP. XII), valorative (CHAP. XIII) et organisationnelle (CHAP. XIV). Leur discussion nous offrira l'occasion de préciser et de nuancer par comparaison notre propre posture. Nous reviendrons aussi, depuis une perspective différente, sur le problème de l'objectivité des fonctions déjà abordé dans la troisième partie.

Première partie :

ANALYSE DU DÉBAT SUR LES FONCTIONS

Modern science is on the whole hostile to teleological explanations. That they are obscurantist and unempirical has been the dominant view among scientists ever since the Renaissance.

Andrew Woodfield, *Teleology* (1976, p. 3)

What role can functions have in a purely scientific description of the world: how can they be "placed" within the framework of current science?

John Bigelow & Robert Pargetter, « Functions » (1987, p. 182)

When we discover a natural function, there are no natural facts discovered beyond the causal facts. Part of what the vocabulary of "functions" adds to the vocabulary of "causes" is a set of values.

John Searle, *The construction of social reality* (1995, p. 15)

Introduction de la première partie

Notre premier objectif dans cette première partie est d'examiner les principales tentatives de naturalisation de la téléologie et des fonctions pour en dégager les éléments essentiels et en identifier les insuffisances et les lacunes les plus importantes. Nous ne prétendons pas réaliser une analyse systématique et exhaustive de l'ensemble des conceptions existantes, mais plutôt une exploration critique de l'évolution du débat depuis les années 1970.

Le second objectif est de voir si les différentes conceptions sont concurrentes et irréconciliables, si elles correspondent à des concepts différents mais homonymes, ou si elles peuvent être conciliées dans une approche commune. Nous faisons l'hypothèse d'une unité sous-jacente à la diversité des formulations.

Le troisième objectif est d'identifier les présupposés ontologiques et épistémologiques communs aux différentes stratégies de naturalisation des fonctions.

La littérature sur les fonctions est très abondante et il n'est pas facile de s'y orienter. Cependant, il semble qu'un consensus se soit imposé autour de deux approches principales : la conception étiologique ou sélectionniste initialement proposée par Larry Wright (1973) et la conception dispositionnelle, dite aussi du « rôle causal », formulée par Robert Cummins (1975). La première vise à expliquer la présence d'un trait à partir de ses effets. La seconde vise à expliquer une capacité d'un système à partir des dispositions de ses parties. Une troisième approche issue de la cybernétique était très répandue entre 1950 et 1970 et connaît un nouvel essor depuis les années 2000. Elle vise à expliquer la téléonomie comme un produit de l'organisation de systèmes plus ou moins complexes. Il faut aussi noter avec Peter Achinstein (1983) qu'au début des années 1980, c'est une approche différente centrée sur les valeurs qui était alors la plus populaire. D'autres, plus ou moins influentes au moment de leur formulation, sont aujourd'hui parfois citées mais rarement discutées. C'est le cas par exemple de l'approche dite mentaliste ou téléomentaliste qui aborde la téléologie depuis une perspective psychologique.

Nous nous limiterons ici à étudier les conceptions étiologiques (diachroniques) et systémiques (synchroniques), ainsi que certaines de leurs combinaisons (approches mixtes). Les approches téléomentaliste, valorative et organisationnelle seront discutées respectivement aux CHAP. XII, XIII & XIV de la cinquième partie.

CHAPITRE I

Approche étiologique (diachronique)

L'approche étiologique a été défendue pour la première fois par Larry Wright dans les années 1970. Elle a été suivie de nombreuses versions et reformulations différentes par différents auteurs. Brièvement, elle consiste à dire que la fonction d'un trait est ce pour quoi il existe, qu'il existe parce qu'il a été sélectionné, et qu'il a été sélectionné en vertu de certains de ses effets[4]. La sélection en question est artificielle (consciente) pour les artefacts et naturelle (darwinienne) pour les organismes biologiques. Par conséquent, attribuer une fonction à un trait, selon la conception étiologique, c'est expliquer son existence en vertu de certains de ses effets.

Certains auteurs adoptent une perspective historique ou tournée vers le passé (*backward-looking*) et analysent la fonction d'un trait comme étant le résultat d'un processus historique de sélection auquel ont contribué certains effets (des ancêtres) de ce trait. D'autres adoptent une perspective tournée vers l'avenir (*forward-looking*) et considèrent que la fonction d'un trait est plutôt sa disposition ou sa propension à contribuer à la sélection présente et future des organismes qui en sont dotés.

Dans la PREMIÈRE SECTION du chapitre, nous nous pencherons sur la formulation de Larry Wright à partir de laquelle se sont développées toutes les versions ultérieures. Dans la SECT. 2, nous étudierons les

4 Allen et Bekoff (1995c, 1995a) proposent les trois composants suivants :

(1) Function claims in biology are intended to explain the existence or maintenance of a trait in a given population;
(2) Biological functions are causally relevant to the existence or maintenance of traits via the mechanism of natural selection;
(3) Functional claims in biology are fully grounded in natural selection and are not derivative of psychological uses of notions such as design, intention and purpose.

formulations historiques de Karen Neander et de Ruth Millikan où la sélection naturelle joue un rôle fondamental. Nous examinerons ensuite, dans la SECT. 3, plusieurs formulations où la « sélection des effets » se conjugue au présent et au futur plutôt qu'au passé. Dans la SECT. 4, nous montrerons d'abord que deux modes d'explication complémentaires se dégagent de ces différentes formulations : une explication causale, qui peut être historique ou pas, et une autre non-causale, qui s'exprime en termes de raisons ou de valeurs ; et nous terminerons avec des questions et objections d'ordre général à propos de l'approche étiologique.

1. Formulation abstraite

1.1. La fonction d'un trait est ce pour quoi il existe

La formulation initiale de l'approche étiologique est due à Larry Wright (1973, 1976).[5] Il s'agit d'une définition générale du concept de fonction qui s'applique aussi bien aux êtres vivants qu'aux artefacts et qui repose sur l'idée que les attributions fonctionnelles, de même que les attributions de buts, sont une forme d'explication : en attribuant un but ou une finalité au comportement d'un agent, on l'explique. De même, attribuer une fonction à un élément d'un système, c'est en expliquer la présence dans le système.[6] Les attributions fonctionnelles sont donc des réponses à des questions de type *pourquoi* :

- « Pourquoi les porcs-épics ont-ils des piquants ? »
- « Pourquoi les vertébrés ont-ils un cœur ? »
- « Pourquoi (certaines) montres ont-elles une trotteuse ? »

Qui plus est, si attribuer une fonction aux piquants des porcs-épics est une manière d'expliquer pourquoi ils en ont, alors la fonction de ces piquants est leur *raison d'être* ; elle est ce pour quoi ils existent :

> « If to specify the function of quills is to explain why porcupines *have* them, then the function must be the reason they *have* them. That is, the ascription of a function must be explanatory in a rather strong sense. » (1973, p. 38)

Par contre, la cause de l'existence ou de la présence de quelque chose n'est pas nécessairement sa fonction. Ainsi, la faculté qu'à l'oxygène de se combiner à l'hémoglobine est la cause de sa présence dans le sang, mais

5 Pour une révision rétrospective de cette formulation, voir Wright (2013).

6 Cette interprétation des attributions fonctionnelles avait été proposée précédemment par Hempel (1965), mais avec des conclusions différentes. Estimant que les explications téléologiques ne s'ajustent pas au modèle déductif-nomologique, Hempel les jugeait invalides.

ce n'est pas sa fonction, car celle-ci consiste essentiellement à apporter de l'énergie à l'organisme par le biais de réactions d'oxydation.

On peut alors, suivant Wright, distinguer deux sortes d'étiologies à partir de la notion de conséquence causale. La faculté de se combiner à l'hémoglobine n'est pas une conséquence de la présence d'oxygène dans le sang, c'est le contraire qui est vrai. En revanche, la production d'énergie est bel et bien une conséquence de la présence d'oxygène dans le sang. Or, la fonction d'un item est toujours une conséquence de sa présence.

La définition étiologique repose donc sur les deux conditions suivantes :

> The function of *X* is *Z means* :
> (1) *X* is there because it does *Z*,
> (2) *Z* is a consequence (or result) of *X*'s being there.

Avec la première condition, Wright inscrit les attributions fonctionnelles dans un schéma explicatif causal (Grene & Depew, 2004, p. 315-316) (voir Fig. 1), mais il insiste sur le fait que l'explication invoquée est causale *au sens large* : « *"because" here is to be taken in its ordinary, conversational, causal-explanatory sense* » (1973, p. 157).

Elle admet, par exemple, que des raisons soient considérées comme des causes. En ce sens, la fonction de *X* n'est pas (ou pas seulement) la cause de son existence, mais plutôt ce que l'on peut appeler avec Wright sa *raison d'être* (qui n'a rien à voir avec le principe de raison suffisante, *nihil est sine ratione*). De plus, cet auteur accorde une importance particulière à la distinction entre explications téléologiques et explications causales au sens strict, et il s'agit d'une distinction entre étiologies, pas entre une étiologie et quelque chose d'autre :

> « functional explanations, although plainly not causal in the usual, restricted sense, do concern how the thing with the function *got there*. Hence they *are* etiological, which is to say "causal" in an extended sense. But this is still a very contentious view. Functional and teleological explanations are usually *contrasted with* causal ones, and we should not abandon that contrast lightly : we should be driven to it. » (1973, p. 156)

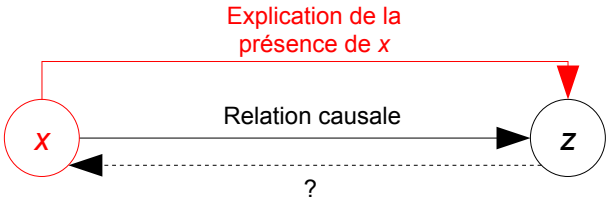

Figure 1: Schéma simplifié de l'approche étiologique de Larry Wright. L'item *X* est présent à cause de *Z* qui est à son tour une conséquence de la présence de *X*.

On peut en effet expliquer pourquoi les porcs-épics ont des piquants et pourquoi les vertébrés ont un cœur en faisant référence à l'organogenèse de l'embryon et aux mécanismes causaux impliqués dans l'expression des gènes correspondants. Et, bien que cette réponse strictement causale soit à la fois correcte et complète, elle est cependant insuffisante ou insatisfaisante, car elle laisse en suspend une autre dimension de la question, à savoir la raison d'être de ces organes. En effet, le même type d'explication causale est également applicable à n'importe quel trait ou item indépendamment de son caractère fonctionnel ou pas. Il est applicable par exemple aux traits vestigiaux dont la présence a toujours une explication causale, mais qui n'ont plus de raison d'être. Il est applicable aussi à des caractéristiques « collatérales » comme le bruit des battements cardiaques et la couleur du sang.

Par ailleurs, le fait que cette première condition laisse la porte ouverte à plusieurs explications différentes permet à la définition d'être applicable indifféremment aux artefacts et aux êtres vivants. En effet, si le cœur des animaux a pour fonction de pomper le sang, peu importe qu'il soit le produit d'une création intelligente ou de la sélection naturelle (L. Wright, 1973, p. 43-44). Dans les deux cas, l'un des effets de l'activité cardiaque explique pourquoi les vertébrés possèdent cet organe. C'est en ce sens que l'on peut dire que pomper le sang est la *raison d'être* du cœur chez les vertébrés.

La première condition n'implique pas une perspective temporelle déterminée, car bien que la forme verbale « *does* » soit au présent, elle est compatible avec une interprétation au passé comme au futur, selon le type d'étiologie pertinente pour expliquer la présence de Z (Neander, 2012a). Cependant, avec la seconde condition, la définition introduit une sorte d'asymétrie temporelle, car elle implique que Z, qui explique l'existence de X, est postérieur à son l'existence.

Ce problème est assez facilement résolu dans le cas d'une création intelligente dans la mesure où le créateur peut envisager les conséquences d'un item X avant que celui-ci n'existe et décider de le fabriquer en raison de celles-ci. Pour les objets naturels, en revanche, à moins de recourir à une causalité rétrograde, les conséquences de X ne peuvent expliquer sa présence ou son existence qu'à un instant t_1 postérieur à celui de sa

première apparition à t_0. La fonction F d'un objet naturel X est donc à la fois une conséquence de l'existence de X à l'instant t_0 et une cause (au sens large) de son existence à l'instant t_1 (voir Fig. 2).

Cela peut s'interpréter de deux manières. La première consiste à dire, par exemple, que chez un individu (particulier) la respiration est une conséquence du fait qu'il est né avec des poumons (t_0) et aussi l'une des

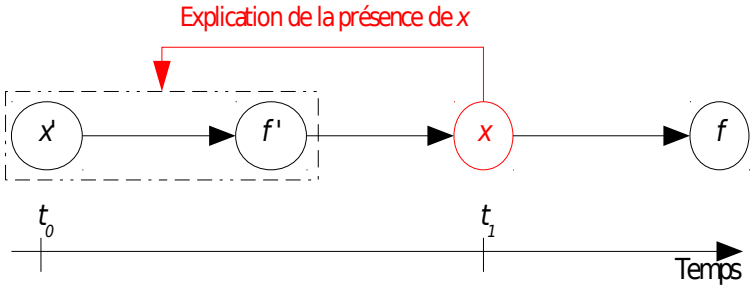

Figure 2: Schéma explicatif de l'approche étiologique de Larry Wright. L'item *X* est présent à l'instant t_1 à cause de *Z'* qui est une conséquence de la présence de *X'* à l'instant t_0.

raisons pour lesquelles il continue à en avoir à t_1 dans la mesure où il est vivant. Mais cette interprétation peut prêter à confusion (L. Wright, 1976, p. 87-88). Il est préférable, en particulier pour les objets biologiques, d'adopter une interprétation où l'item X fait référence à un type plutôt qu'à *token* (L. Wright, 1976, p. 113) et où, par conséquent, la présence de poumons chez un individu (quelconque) s'explique par le fait que ce type d'organe permet aux individus (en général) de respirer.

L'interprétation de la définition en termes de types rend compte de la dimension normative que nous reconnaissons intuitivement au concept (S. D. Mitchell, 1995, p. 45; Proust, 1995). Elle permet de dire qu'un cœur malformé, incapable de pomper le sang, n'en conserve pas moins la fonction qui correspond à son type. La fonction du cœur est une conséquence typique, donc *censée* se produire, même quand *de fait* un cœur particulier en est incapable (L. Wright, 1976, p. 112).

1.2. Contre-exemples et solutions

La définition de Wright prête le flanc à plusieurs contre-exemples. Imaginons qu'une fuite dans un tuyau de gaz toxique provoque la mort de ceux qui essaient de la réparer[7]. La fuite est une conséquence de l'existence d'un trou dans le tuyau, et elle explique pourquoi le trou continue

7 Ce contre-exemple a été formulé par Christopher Boorse (1976, p. 72) et repris par d'autres (Kitcher, 1993, p. 264). Bedau (1991, p. 648) propose un exemple analogue avec un objet naturel non vivant : un bâton dans un cours d'eau.

à être présent. Conformément à la définition, il faudrait dire que la fonction du trou est de laisser s'échapper le gaz toxique. Pourtant, cette attribution fonctionnelle est contre-intuitive. On ne peut pas dire que le trou soit là *pour* laisser s'échapper le gaz, à moins qu'il n'ait été percé intentionnellement dans ce but.

Le problème ne vient pas du fait que le tuyau soit un objet artificiel, car le même type de contre-exemple peut facilement être reproduit avec des objets naturels. Par exemple, le virus Ebola produit une protéine (VP24) qui empêche l'activation précoce du système immunitaire et lui permet de se développer en grande quantité avant que celui-ci n'intervienne pour le détruire (Xu et al., 2014). De cette manière, il assure sa présence dans l'organisme parce qu'il bloque le système immunitaire (condition 1) et ce blocage est une conséquence de l'existence du virus (condition 2). On devrait donc pouvoir dire, conformément à la définition de Wright, que l'une des fonctions du virus Ebola est de bloquer la réponse immunitaire de l'organisme hôte. Pourtant, cette attribution fonctionnelle est contre-intuitive[8]. Autre exemple : l'excès de cholestérol forme des plaques d'athérome sur la paroi interne des artères (athérosclérose). Ces plaques entraînent l'occlusion des artères non seulement par leur seule présence mais aussi et surtout par les conséquences de cette présence (recouvrement par d'autres molécules et cellules, durcissement, formation de caillots). Les deux conditions de la définition étiologique étant satisfaites, on devrait pouvoir dire que les dépôts de cholestérol ont pour fonction de boucher les artères, ce qui n'est pas le cas.

De manière générale, la définition est satisfaite dès lors que se met en place une rétro-alimentation positive (l'équivalent d'une auto-sélection), indépendamment de la nature de l'item et du caractère bénéfique ou délétère de ses conséquences.

On trouve décrites dans la littérature spécialisée trois manières de lever ces contre-exemples. La première consiste à éliminer de la définition la condition (2) et à limiter les attributions fonctionnelles aux items qui sont le fruit d'une histoire sélective. Autrement dit, l'item X a la fonction Z si et seulement si X est présent parce qu'il a été sélectionné pour faire Z. C'est la solution qu'adoptent les partisans d'une définition en termes de sélection naturelle (voir SECT. 2 ci-après).

La seconde consiste à rappeler que chez Wright l'explication de la présence d'un item doit être interprétée à la lumière du type auquel cet item appartient[9]. Mais cette solution ne semble pas en mesure de

8 On peut dire que la protéine VP24 a une fonction *pour le virus*, puisque le blocage de la réponse immunitaire lui permet de se développer impunément dans l'organisme, mais on peut difficilement dire qu'elle en ait une pour l'hôte qui souffre l'infection et finit souvent par en mourir.

9 Cette solution est notamment avancée par Zellner (2001, p. 506) : « *At least in the actual world there is no reason why the fact that the gas typically asphyxi-*

répondre à des contre-exemples comme celui du virus et du cholestérol que nous avons exposés plus haut.

La troisième consiste à limiter la définition aux conséquences d'un item dont on peut dire qu'elles sont bénéfiques ou avantageuses, ce qui n'est évidemment pas le cas des exemples précédents, à moins de considérer que la fuite de gaz soit un acte de malveillance qui, d'une manière ou d'une autre, bénéficie à celui qui l'a intentionnellement perpétré[10]. Cette troisième solution est privilégiée par les partisans de l'approche valorative des fonctions, comme Ayala (1970), Woodfield (1976), Bedau (1992b) et McLaughlin (2001). Nous y reviendrons au CHAP. XIII.

Aux trois précédentes, nous ajoutons quant à nous une quatrième solution qui s'inspire directement de l'approche organisationnelle des fonctions (CHAP. XIV). Elle consiste à interpréter la première condition de la définition de Wright de sorte que la présence de X ne s'explique pas directement par ses effets Z, mais par l'intermédiaire d'un tout organisé dont X (ou plutôt sa fonction) est une partie. C'est-à-dire par exemple que Z contribue à l'existence d'un système S qui à son tour contribue à la présence de X. Ainsi, on pourrait dire que le cœur a pour fonction de pomper le sang, car cette activité contribue à la survie de l'organisme, lequel contribue à son tour au maintient du cœur, par exemple sous la forme du renouvellement cellulaire. Il n'y a pas de système équivalent dans le cas du trou dans le tuyau de gaz toxique, ni dans celui des plaques d'athérome, de sorte que l'on ne pourrait pas leur attribuer de fonction.

1.3. La conception de Wright n'est pas indifférente à l'origine causale

Examinons maintenant la question de la généralité. Wright affirme que sa définition est applicable aussi bien aux artefacts qu'aux organismes biologiques. Il prétend, de cette manière, évacuer le problème de l'intentionnalité en montrant qu'il n'y a pas de vraie différence entre les

---

ated scientists would cause the presence of the gas in individual instances. The gas is present in a given case because it asphyxiated this scientist, not because the presence of the gas typically asphyxiates scientists. »

10 L'analyse de Wright laissait volontairement de côté cet aspect des choses, car selon lui une conséquence utile n'est pas forcément fonctionnelle et une fonction n'est pas forcément bénéfique ou avantageuse. En effet, d'après lui, la fonction d'un item n'est pas, ou pas seulement, ce pour quoi cet item est bon (*what it is good for*), et la question « *Pourquoi les animaux ont-ils un foie ?* » ne peut pas être rendue par « *À quoi le foie est-il bon ?* », car le foie peut être bon pour beaucoup de choses qui ne sont pas sa fonction, y compris bon pour dîner avec des oignons (1973, p. 155-156). Cependant, il intégrait la valeur d'un item dans le cadre d'une explication de sa présence : « *the* function *of the liver is that* particular *thing it is good for which explains why animals have them* » (1973, p. 156).

fonctions dites « conscientes » et les fonctions naturelles. Tout au plus reconnaît-il que l'interprétation en termes de types donne lieu à une légère différence de sens, car la sélection consciente, contrairement à la sélection naturelle, peut opérer sur un *token* et ne nécessite pas plusieurs générations d'individus (L. Wright, 1976, p. 89). Cette différence de sens est la seule qu'il concède. Pourtant, on peut montrer qu'elle en implique d'autres.

La première d'entre elles concerne la portée de l'explication. Les fonctions conscientes expliquent ou peuvent expliquer l'existence d'un item à partir de sa première apparition (t_0) tandis que les fonctions naturelles ne deviennent explicatives qu'à partir d'un instant postérieur (t_1), lequel correspond — dans le cadre de la sélection naturelle — à une génération postérieure. Quand on attribue au cœur la fonction étiologique de pomper le sang, cette attribution n'a donc pas le même sens selon que cet organe est une création divine ou un produit de la nature. Dans le premier cas, on peut dire que tous les cœurs — y compris ceux d'Adam et Eve — possèdent la fonction en question et sont explicables par elle[11]. Dans le second cas, en revanche, la portée de l'explication se limite aux descendants du ou des premiers individus, de telle sorte que les cœurs d'Adam et d'Eve (et peut-être ceux de leurs enfants) n'ont aucune fonction. Suivant la définition étiologique de Wright, un même cœur particulier — celui d'Adam — aura ou pas une fonction selon qu'il est un objet naturel ou artificiel, alors même que la définition se veut également applicable aux deux étiologies. Cela semble également vouloir dire que le cœur d'Adam ou d'Ève n'est fonctionnellement explicable qu'en référence aux intentions d'un créateur ou, par exemple, à une finalité naturelle ayant guidé son apparition.

Une autre différence apparaît à travers l'usage du concept de sélection. Pour justifier la fonction des organes dans un organisme, Wright déclare que la sélection naturelle agit de manière *analogue* à la sélection artificielle. Dans les deux cas, explique l'auteur, il s'agit du même type de sélection. L'objectif de cette analogie est de montrer que les fonctions des artefacts ne sont pas plus relatives à une intention consciente que celles des organismes biologiques. Pour ce faire, il distingue deux types de sélection artificielle : la simple « discrimination » et la « sélection des conséquences » (*consequence-selection*). La première correspond à un choix non motivé, comme lorsque l'on choisit sans raison particulière un objet parmi d'autres placés devant nous. L'autre correspond à un choix motivé par un avantage particulier que procure la chose sélectionnée : « Je choisis X parce qu'il fait Z ». Or, selon Wright, la sélection naturelle correspond

11 Il arrive aussi que les innovations techniques soient le fruit d'un heureux hasard et qu'une fonction nouvelle apparaisse suite à une conséquence imprévue, mais ce cas de figure, contrairement à l'évolution par sélection naturelle, n'est pas la règle.

à ce second type de sélection parce que, dit-il, la volition y joue un rôle marginal, de sorte que l'on peut parler d'auto-sélection[12] :

> « We might want to say that *natural* selection is really *self*-selection, that nothing is *doing* the selecting; given the nature of X, Z, and the environment, X will *automatically* be selected. » (L. Wright, 1976, p. 86)

> « Given our criteria, we might well say that X *does* select itself in conscious consequence-selection. By the very nature of X, Z, and our criteria (the implementation of which may be considered the environment), X will automatically be selected. » (L. Wright, 1973, p. 163)

On peut exprimer la même distinction entre ces deux types de sélection en termes de choix rationnel et de choix arbitraire. À la fin du passage cité, l'auteur ajoute une note de bas de page disant : « *c'est une version du vieux problème de la tension entre rationalité et liberté.* » En effet, dans un choix rationnel, on peut dire que la meilleure option s'impose d'elle-même. De cette manière, Wright veut donc montrer que le caractère conscient ou pas de la sélection est secondaire face au mécanisme d'auto-sélection qui opère aussi bien dans la sélection naturelle que dans l'artificielle. Il veut ainsi gommer toute référence explicite à une intention ou à un but pour les fonctions conscientes (1973, p. 164-165). Mais l'auteur se garde bien de souligner que le type de sélection rationnelle qu'il invoque est implicitement relatif au choix d'une finalité (ou au choix des critères que mentionne l'auteur). Si je choisis X c'est parce que, en vertu de ses caractéristiques ou de ses conséquences Z, il représente le meilleur moyen pour atteindre une certaine fin. Donc, s'il y a quelque chose comme une auto-sélection de X par ses conséquences Z, elle est relative au choix d'une fin qui, pour les fonctions conscientes, est plus ou moins arbitraire. C'est-à-dire que la sélection des conséquences dépend malgré tout d'un autre type de sélection où la volition est directement présente.

Une troisième différence est liée aux erreurs de sélection. Imaginons que quelqu'un construise une maison pour hirondelles et que, par erreur, il la fabrique avec une ouverture tellement petite que les oiseaux ne puissent ni entrer ni sortir (Nissen, 1997, p. 147). La fonction de X (l'ouverture) est Z (permettre aux hirondelles de rentrer et sortir), bien qu'il soit impossible que X ait la conséquence Z. À cela, Wright (1976, p. 113) répond que X est supposé avoir la conséquence Z parce que *quelqu'un* a supposé que X permettrait Z, et *c'est là* la raison pour laquelle X est présent. L'étiologie, ajoute-t-il, est la même que si la chose avait fonc-

12 Cette interprétation de la sélection naturelle est critiquable. Voir par exemple Nissen (1997, p. 144).

tionné, sauf qu'elle ne fonctionne pas. Pourtant, l'explication fait directement appel à l'intentionnalité d'un créateur alors que celle-ci n'a pas de place dans la définition. Encore une fois, il est possible de contourner le problème en disant que les structures de ce type dans les maisons pour oiseaux ont la fonction de laisser entrer et sortir les volatiles et que, par conséquent, l'ouverture de cette maison en particulier possède bel et bien la fonction correspondant à son type. Mais cette réponse ne s'applique qu'aux objets appartenant à un type pré-existant et pas aux objets nouveaux. Dans ce dernier cas, il semble que la fonctionnalité d'un artefact défectueux soit relative aux intentions de son créateur.

2. Formulations historiques

Les principales réponses aux contre-exemples soulevés par la définition étiologique de Wright consistent à limiter les attributions fonctionnelles à un type d'histoires causales : celles où opère une sélection des effets ou des conséquences. Bien qu'elle fut importante pour lui, l'idée de sélection (consciente ou naturelle) n'était pas explicitement mentionnée dans la définition. C'est la raison pour laquelle un trou accidentel dans un tuyau de gaz pouvait se voir assigner une fonction. L'idée centrale commune aux conceptions dites « sélectionnistes » (*selected effects*) est qu'un item X ne peut avoir la fonction F que s'il a été sélectionné pour cela, ou plutôt *à cause* de cela. David Buller en propose la formulation standard suivante :

> « A current token of a trait *T* has the function of producing an effect of type *E* just in case, at some point in evolutionary history, there was selection for *T* (over alternative items) because of its having produced effects of type *E*. » (Buller, 1998)

Cette conception sélectionniste a été principalement développée par Ruth Millikan (1984, 1989b) et Karen Neander (1991b, 1991a). Chez la première, la définition initiale du concept de « fonction propre » était liée à un projet de naturalisation de l'intentionnalité et visait à rendre compte des caractéristiques communes à différentes catégories de choses : non seulement les objets biologiques, mais aussi les comportements, les dispositifs du langage (mots, phrases), les états mentaux, les artefacts, etc. Bien que développée indépendamment de celle de Wright, sa définition en reprenait l'idée centrale, à savoir qu'un item a une fonction si sa présence ou ses caractéristiques s'expliquent par l'une de ses conséquences. De manière à éviter toute circularité ou causalité rétrograde, il fallait que les conséquences responsables de la présence de l'item fonctionnel X fussent imputables non pas à l'item lui-même mais à l'un de ses prédécesseurs de même type, le second item étant une reproduction ou une copie du premier en ce qui concerne les caractéristiques fonctionnelles. À cela

s'ajoutait une autre condition visant à exclure les effets du premier item pouvant contribuer seulement *par accident* à la présence du second. Cette seconde condition est la sélection : un effet de l'item X ne peut être considéré comme sa fonction que si X a été *sélectionné*, *parmi d'autres* items, *parce qu*'il avait cet effet (1993, p. 35).

La définition simplifiée que propose Millikan est la suivante :

> [U]n item X a la fonction propre F si X provient de la "reproduction" (par exemple comme une copie, ou une copie de copie) d'un ou de plusieurs items antérieurs qui ont accompli F dans le passé *parce qu*'ils possédaient les propriétés reproduites (Millikan, 1989b, p. 288 [traduction libre]).

Cette définition repose sur un certain nombre de concepts, comme ceux de « reproduction » et de « famille reproductivement établie », dont la clarification requerrait une analyse minutieuse (voir Millikan, 1984).

Karen Neander (1991b) est plus directe en affirmant que la fonction propre d'un trait est l'effet pour lequel ce trait a été sélectionné. En ce qui concerne les fonctions biologiques, il s'agit de la sélection naturelle dans le cadre de la théorie darwinienne de l'évolution. Et puisque la sélection naturelle opère sur des populations entières plutôt que sur des individus, ce ne sont pas des *tokens* qui sont sélectionnées mais des types. Comme chez Wright et Millikan, la fonction propre d'un *token* est relative à celle du type auquel il appartient, car c'est ce type de trait qui a été historiquement sélectionné. Du point de vue biologique, cela implique que le trait en question est héréditaire et que la fonction d'un *token* de ce trait est liée à ses effets chez ses ancêtres. En prenant le génotype comme unité de sélection, Neander définit les fonctions biologiques comme suit :

> It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X's* type did to contribute to the inclusive fitness of *O's* ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection. (Neander, 1991a, p. 174).

En ce qui concerne les artefacts, elle affirme que la sélection naturelle n'est pas applicable et que leurs fonctions sont relatives aux intentions ou aux fins pour lesquels ils ont été inventés, créés ou utilisés par un agent (1991b, p. 462). Il n'en demeure pas moins que leur fonction est l'effet pour lequel ils ont été (intentionnellement) sélectionnés. Selon cette auteure, la principale différence entre les fonctions biologiques et techniques réside, d'une part, dans le fait que ces dernières ne

sont pas nécessairement relatives à des types (puisqu'une invention unique et originale peut néanmoins être dotée de fonctions), et, d'autre part, dans le fait que les effets d'un artefact pouvant être anticipés mentalement par un agent il n'est pas nécessaire de faire appel à des « ancêtres » ayant eu ces effets par le passé. À vrai dire, dit-elle (1991a, p. 175), il n'est pas même nécessaire que l'effet en question soit jamais réalisé (dans le passé, au présent, ou dans le futur).

Une caractéristique importante commune aux propositions de Millikan et de Neander est leur défense de la normativité des fonctions propres et la relation qu'elle entretient avec la question des catégories biologiques. Le cœur, les poumons et la plupart des traits organiques désignent des types de structures qui peuvent être morphologiquement très différentes les unes des autres, notamment d'une espèce à l'autre. Ces structures diverses sont rangées dans des catégories biologiques communes en vertu, non pas de ce qu'elles font, mais de ce qu'elles sont *censées* faire, c'est-à-dire en vertu de la fonction qu'elles accomplissent. Un cœur malade ou sévèrement malformé, incapable de pomper le sang, est néanmoins censé accomplir cette fonction dans la mesure où il appartient à la catégorie d'organe correspondante, celle des « cœurs ». La clarification du concept de fonction propre d'un trait (ce qu'il est censé faire) est donc liée au problème de sa catégorisation[13]. Or, ce qu'un trait est censé faire est ce pour quoi il a été sélectionné.

Trois objections classiques ont été formulées à l'encontre des conceptions étiologiques basées sur la sélection naturelle (Bigelow & Pargetter, 1987; Boorse, 1976; Nagel, 1977a, p. 284; Nissen, 1997; L. Wright, 1976). La première dit qu'elle est historiquement fausse et que l'on peut attribuer une fonction à un organe sans avoir la moindre idée de son étiologie. Ainsi, lorsque Harvey découvrit la fonction du cœur en 1616, il ne connaissait pas la théorie darwinienne et ne pouvait donc pas penser que la circulation du sang était la cause pour laquelle les cœurs

13 Voir à ce propos la discussion entre, d'un côté, Robert Cummins (2002b) et Peter McLaughlin (2001) qui nient qu'il y ait eu sélection des types fonctionnels, comme le cœur pour la circulation du sang ou les ailes pour le vol, d'un autre côté, Ron Amundson & George Lauder (1994), Paul Griffiths (1994, 2006) et Paul Sheldon Davies (2001), qui nient le rôle des fonctions dans les catégories biologiques, si ce n'est dans les catégories analogues, et, du côté diamétralement opposé, Alex Rosenberg et Karen Neander (Neander, 2002, 2010; Neander & Rosenberg, 2012; Rosenberg & Neander, 2009), qui revendiquent au contraire l'existence de catégories homologues-fonctionnelles, lesquelles incluent des traits homologues ayant la même fonction étiologique, comme les cœurs des vertébrés et les ailes des oiseaux. Ajoutons à cela la polémique issue de l'objection de Nanay (2010) : si la fonction d'un trait est déterminée par son type et si les types sont déterminés par leurs propriétés fonctionnelles, alors il y a un cercle.

avaient été favorisés par la sélection naturelle. Par conséquent, la théorie étiologique ne rendrait pas compte de l'usage que font les biologistes du concept de fonction. La réponse de Millikan (1989b) à cette objection consiste à dire que sa définition est purement théorique et ne vise pas à rendre compte de la manière dont les biologistes emploient le concept. Neander (1991a, p. 176) répond quant à elle en disant qu'elle ne s'intéresse pas à l'histoire du concept de fonction mais à son usage contemporain et que les concepts scientifiques évoluent avec le temps et avec les théories sur lesquelles ils s'appuient. Cette réponse laisse de côté le fait que tous les biologistes actuels n'emploient pas forcément le concept de la même façon et qu'il est manifestement possible de le faire sans référence aucune à la sélection naturelle.

La seconde objection classique affirme que cette conception est analytiquement arrogante, car elle écarte la possibilité que les êtres vivants soient le fruit d'une création intelligente. En d'autres termes, sa validité est directement liée à celle de la théorie darwinienne. La réponse de Neander (1991a, p. 177-178) consiste à dire que la vérité de la définition n'est pas empirique mais conceptuelle et qu'elle est relative à une communauté linguistique à un moment donné, de sorte que : ou bien la théorie de la sélection naturelle est vraie, ou bien il n'y a pas de fonctions propres *au sens où l'emploient les biologistes contemporains*. La conception étiologique entendue comme une définition théorique serait fausse si la théorie de la sélection naturelle s'avérait fausse. En revanche, entendue comme une analyse conceptuelle, sa vérité ne dépend pas de celle de la théorie évolutionniste, mais seulement de l'usage du concept par les biologistes.

La troisième objection est qu'il semble contre-intuitif qu'un organisme soit dépourvu de fonctions s'il n'a pas d'histoire. Par exemple si un accident cosmique faisait apparaître des créatures en tous points identiques à des lions (Boorse, 1976, p. 74), ou si le monde n'était âgé que de cinq minutes, selon l'hypothèse de Russell (Bigelow & Pargetter, 1987, p. 188), ou si Dieu avait effectivement créé Adam et Eve de manière instantanée (Plantinga, 1993, p. 203). Chacune de ces expériences de pensée nous met face à des êtres vivants qui n'ont pas d'histoire sélective et dont les traits, par conséquent, n'ont pas de fonctions — bien qu'ils soient par ailleurs indistinguables de ceux auxquels nous attribuons des fonctions. Ces contre-exemples sont analogues à celui de « l'homme du marais » (*Swampman*), formulé par Davidson (1987) contre la téléosémantique, qui a suscité une vaste littérature dans laquelle nous n'allons pas entrer.

Dans le monde réel, le même problème se pose lorsque l'on se penche sur les origines de la vie ou sur l'apparition de nouvelles fonctions. L'objection a été formulée en termes d'épiphénoménalisme (Saborido, 2012), c'est-à-dire que l'attribution de fonctions n'apporte aucune information additionnelle quant au « phénomène » à analyser, à savoir le trait, car elle ne dépend ni de sa structure, ni de ses propriétés, ni de ses effets, ni de sa place ni de son rôle actuels dans l'organisme.

Karen Neander (1991a, p. 179-180) prend la mesure de l'objection, mais la manière dont elle y répond finalement est peu convaincante. Supposons, dit-elle, que les lions n'existent pas et que, soudain, apparaisse une demi-douzaine de lions on ne sait pas comment. Après avoir observé avec étonnement ces curieux animaux, on s'interroge sur les protubérances ressemblant à des ailes qu'ils ont de chaque côté du corps. Ces membres ont-ils la fonction propre de voler ? On constate que les lions ne peuvent pas voler, mais ces protubérances pourraient être des ailes mal formées, malades ou atrophiées. Ou peut-être que les lions ne peuvent pas voler parce qu'ils se trouvent en dehors de leur habitat naturel où le champ gravitationnel est plus faible. D'un autre côté, il pourrait également s'agir de structures vestigiales, donc dénuées de fonction. L'argument de Neander consiste à dire qu'on ne peut pas classer les membres en question dans une catégorie d'organes tant qu'on ne connaît ou qu'on n'infère pas leur histoire. Et s'ils n'ont pas d'histoire, s'ils sont le fruit d'un accident cosmique, alors ils n'ont pas de fonction et ne peuvent ni dysfonctionner ni avoir perdu leur fonction.

Sauf erreur d'interprétation de notre part, la réponse de Neander est fort peu raisonnable, car il suffirait de quelques minutes d'observation pour se convaincre que les protubérances latérales de ces curieux animaux sont faites pour marcher, courir, attraper, etc. Parce que c'est ce qu'elles font et parce qu'elles le font très bien. En revanche, il n'y a aucune raison apparente pour penser qu'il pourrait s'agir d'ailes mal formées ou de structures vestigiales. Cette hypothèse nous semble donc absurde ou singulièrement mauvaise[14]. Il n'est d'ailleurs pas nécessaire d'imaginer un accident cosmique pour s'en rendre compte. Il suffit de se confronter à n'importe quel être vivant jusqu'ici inconnu, qu'il soit actuel ou fossile, terrestre ou extraterrestre. À première vue, les attributions fonctionnelles sont basées sur l'observation des traits organiques, de leurs effets éventuels, de leur possible rôle causal dans l'organisme et des similitudes structurelles ou fonctionnelles avec d'autres organismes connus. C'est à partir de cette observation que l'on peut ensuite formuler des hypothèses quant à leur étiologie. Il en va de même pour les artefacts. C'est à partir de l'observation des caractéristiques et du fonctionnement (effets, comportement) d'un objet que l'on peut inférer la fonction de ses parties et, partant de là, son histoire causale[15]. Contrairement à ce que semble

14 À moins d'imaginer que nous nous trouvions sur une planète (par exemple une planète gazeuse) où tous les animaux ont des ailes au lieu de pattes, c'est-à-dire sur une planète où cette catégorie d'organes n'existe pas.

15 La compréhension du fonctionnement des parties d'un organisme (ou d'un artefact) est la meilleure — et parfois la seule — manière de comprendre son étiologie, comme l'indique Michael Ruse : « *Rather than being thrown back at once on impossible questions about how natural selection operated in the distant past, evolutionists begin by asking how things work at the moment. When*

défendre Neander, la connaissance de l'histoire causale est secondaire par rapport au classement dans une catégorie fonctionnelle, au moins tant qu'on se limite à la formulation d'hypothèses.

Millikan (1989b, p. 292-293) refuse pour sa part de considérer les hypothèses logiquement possibles, mais irréelles. Selon sa définition, un double accidentel n'a pas de fonctions parce qu'il n'a pas l'histoire qu'il faut : il n'est en effet ni la reproduction ni le produit d'un objet doté de fonctions. Sa réponse à ce type d'objections consiste à dire qu'une définition théorique n'a pas à rendre compte de situations qui, bien que logiquement possibles, sont de fait inexistantes. Et puisque ces situations n'existent pas, alors la présence de certaines propriétés et dispositions dans un organisme est un critère infaillible de la possession des fonctions correspondantes, mais il ne faut pas confondre les uns et les autres ; ce n'est pas le fait d'avoir certaines propriétés et dispositions qui constitue le fait d'avoir une fonction (1989b, p. 293).

D'autres critiques formulées par Peter McLaughlin (2001) remettent en question l'idée centrale de l'approche étiologique selon laquelle les traits biologiques reçoivent leur fonction de la sélection. Tout d'abord, il est discutable que la nature ait pu sélectionner les cœurs pour pomper le sang et pas pour faire du bruit, car il est douteux que la nature ait pu choisir entre des animaux à cœur bruyant et d'autres à cœur silencieux. Autrement dit, à défaut de variations en ce sens, il est douteux que la sélection naturelle ait pu distinguer entre ces deux propriétés du cœur comme nous le faisons, car elles sont étroitement liées.

De plus, l'approche étiologique conçoit la sélection naturelle sur le modèle de l'artefact, où l'horloger sélectionne les rouages et les ressorts en vertu de leurs fonctions et les assemble pour en faire une montre. Le problème, dit McLaughlin, est que la nature n'agit pas de cette façon. Elle ressemblerait plutôt à un éleveur de pigeons ou de moutons qui ne peut pas sélectionner directement des traits, mais seulement des organismes entiers. Or, les organismes entiers n'ont pas de fonctions. L'horloger sélectionne et manipule des rouages et des ressorts, c'est-à-dire des éléments internes, dans le but d'altérer les propriétés de la montre. L'éleveur, en revanche, doit sélectionner des organismes entiers et manipuler les conditions extérieures (par le contrôle de la reproduction, notamment) pour altérer les propriétés des éléments internes. Par conséquent, on ne peut pas dire que la fonction d'un trait soit ce pour quoi il a été sélectionné, car les traits ne sont pas sélectionnés, tandis que les organismes entiers — qui eux le sont — n'ont pas de fonctions.

---

once they have got a handle on this question, they are able next to talk in terms of natural selection, trying to relate their studies back to what happened in the past. But this second stage can occur only if first they have used their design metaphor to ask pertinent questions about function. » (Ruse, 2002, p. 47)

Une autre objection formulée par Bence Nanay (2010) dit que la conception étiologique est circulaire, car la fonction d'un trait est relative aux propriétés des *tokens* du même type, alors que les types sont eux-mêmes déjà déterminés (en partie) par leurs propriétés fonctionnelles. Un argument similaire avait été formulé précédemment par Davies (2001). L'objection est discutée notamment par Kiritani (2011), Neander (2010), Neander & Rosenberg (2012).

3. Formulations propensionnistes

Les définitions étiologiques basées sur la sélection naturelle consistent généralement à dire que la fonction d'un item ou d'un trait explique causalement son existence ou sa conservation[16] dans une population donnée par l'intermédiaire du mécanisme de sélection naturelle. Ces définitions se distinguent notamment par la perspective temporelle adoptée. Certains auteurs adoptent une perspective tournée vers le passé (*backward-looking*) qui fait dépendre les fonctions d'une histoire sélective plus ou moins longue, comme Neander et Millikan, ou d'une histoire récente, comme Griffiths (1993), Godfrey-Smith (1994) et Schwartz (2002). D'autres, comme Ruse (1971, 1973a), Bigelow et Pargetter (1987) ou Walsh (1996) adoptent une perspective tournée vers l'avenir (*forward-looking*) et interprètent le concept de fonction en termes dispositionnels. La différence n'est pas anodine, car cette seconde manière d'aborder le problème semble apporter une réponse aux trois objections précédentes :

1. On peut attribuer une fonction à un organe sans avoir la moindre idée de son étiologie (exemple de Harvey).
2. Les définitions historiques fondées sur la sélection naturelle reposent sur des bases contingentes, car si la théorie darwinienne s'avérait fausse il n'y aurait pas de fonctions biologiques.
3. Il semble contre-intuitif qu'un organisme soit dénué de fonctions parce qu'il n'a pas d'histoire.

L'un des premiers auteurs à défendre une conception tournée vers l'avenir est Michael Ruse (1971, 1973a), pour qui les énoncés fonctionnels n'ont de sens que dans le contexte de la théorie de l'évolution et doivent être compris en termes d'adaptation, c'est-à-dire de contribution à la survie et au succès reproductif de l'organisme[17] ; mais contrairement à Wright (1972), qui met l'accent sur une adaptation effective (avérée

16 Certains auteurs distinguent l'apparition et la diffusion d'un trait dans une population de la préservation de celui-ci après son apparition (Allen, 2003; Purton, 1979).

historiquement), Ruse adopte une perspective dispositionnelle qui l'autorise à récuser les réticences de Wright à attribuer une fonction aux premières générations d'un item alors qu'on peut le faire pour ses descendants (Ruse, 1973a, p. 279). Une adaptation, en premier lieu, selon Ruse, est quelque chose qui augmente les chances de survie et de reproduction de certains organismes par rapport à d'autres (par ailleurs identiques) qui en sont dépourvus[18]. Les fonctions et les attributions fonctionnelles ne dépendent donc pas de l'étiologie de l'item en question mais de l'analyse comparative des capacités de survie et de reproduction que la présence ou l'absence de cet item confère aux organismes, toutes choses étant égales par ailleurs.

Il propose la définition suivante :

> Dire qu'un item *X* dans le système *S* a la fonction *F* signifie[19] :
> i. *S* réalise (ou peut réaliser) *F* au moyen de *X* ;
> ii. la réalisation de *F* est adaptative pour S ; et
> iii. *X* est une adaptation (pour la réalisation de *F*). »

Bien que la proposition de Michael Ruse se limite aux organismes biologiques dans le cadre de la théorie de l'évolution, on peut envisager sa généralisation. Imaginons par exemple que les adaptations, au lieu d'être causées par un mécanisme naturel, soient en réalité le résultat d'une intervention intelligente (humaine, divine ou extraterrestre, peu importe). Du point de vue de l'organisme lui-même comme du point de vue d'un observateur extérieur rien ne serait changé. Grâce à une modification de leur physiologie ou de leur comportement, certains organismes s'avéreraient mieux adaptés que leurs semblables à leur environnement commun et auraient donc des chances de survie et de reproduction supérieures. On pourrait ainsi leur attribuer des fonctions, définies de manière analogue à celles de Ruse, dans un cadre qui n'est pas strictement darwi-

17 Robert Brandon (1981) dit à peu près la même chose que Ruse (à propos du langage téléologique de la biologie et de sa relation avec les adaptations entendues de façon propensionniste) tout en évitant intentionnellement d'employer le concept de fonction.

18 Un trait Y est une adaptation, selon Ruse (1973a, p. 278), si et seulement si (a) les organismes qui en sont dotés ont une plus grande disposition à survivre et à se reproduire que des organismes identiques si ce n'est par le fait qu'ils sont dépourvus de Y ; et (b) les organismes (passés, présents ou futurs) dotés de Y survivent et se reproduisent davantage — en partie grâce à Y — que les organismes qui en sont dépourvus. Plus récemment, Reeve et Sherman (1993) ont également défendu une définition non-historique du concept d'adaptation, contrairement à Sober (1984, p. 208).

19 Ruse (1982, p. 304), cité par McLaughlin (2001, p. 87).

nien. On pourrait aller encore plus loin en interprétant le concept d'adaptation en termes non biologiques et en l'appliquant aux artefacts[20]. On peut d'ailleurs concevoir les fonctions techniques en termes de sélection socioculturelle (Longy, 2009).

De notre point de vue, la différence la plus intéressante entre les propositions précédentes et celle de Ruse est celle que soulignent Bigelow et Pargetter (1987, p. 189) à propos du concept de *fitness.* Les partisans d'une approche historique estiment que la *fitness* (entendue comme succès reproductif ou reproduction différentielle) ne peut être jugée que rétrospectivement. Les partisans d'une approche dispositionnelle, de l'autre côté, considéreraient que la *fitness* (entendue comme valeur adaptative) est une propriété dispositionnelle de l'organisme (ou de l'espèce) dans un environnement donné. Pour les premiers, un organisme serait en quelque sorte « meilleur » que ses semblables dans la mesure où il a survécu et a eu une descendance plus nombreuse ; pour les seconds, si un organisme survit et se reproduit davantage c'est parce qu'il doit être « meilleur » — parce qu'il y a quelque chose chez lui, dans sa structure ou son comportement, qui augmente ses chances[21].

---

20 En reprenant mot pour mot la définition de Ruse (1973:277-278), on devrait pouvoir remplacer un énoncé fonctionnel de la forme :

> « La fonction du modem [X] dans un ordinateur [Z] est de lui permettre de communiquer avec d'autres ordinateurs distants [Y] »

par les deux énoncés suivants :

> (i) l'ordinateur [Z] communique [Y] en utilisant le modem [X] ;
>
> (ii) la communication entre ordinateurs [Y] est une adaptation.

où quelque chose, [Y], est une adaptation ssi

> (a) les ordinateurs [Z's] capables de [Y] ont plus de chances de survivre et de se reproduire (métaphoriquement) que des ordinateurs [Z] identiques si ce n'est par le fait qu'ils sont incapables de [Y].
>
> (b) il y a eu, il y a ou il y aura effectivement un certain nombre d'ordinateurs capables de [Y] qui survivent et se reproduisent (métaphoriquement) à cause de cette capacité [Y], tandis que d'autres ordinateurs — identiques si ce n'est par le fait qu'ils sont incapables de [Y] — ne survivent pas et ne se reproduisent pas (en conséquence de quoi les items X qui rendent possible [Y] sont préservés par sélection parmi la population des ordinateurs [Z's]."

21 Si la *fitness* dépend seulement de la survie effective, alors rien n'empêche qu'elle soit due au hasard, à une série de coups de chance : « *Consequently, what confers a function is not the sheer fact of survival-due-to-a-character, but rather, survival due to the propensities the character bestows upon the creature.* » (Bigelow & Pargetter, 1987, p. 192). Philip Kitcher (1993, p. 269) soulève ce problème en se demandant si une attribution fonctionnelle sur la base de la sélection naturelle implique que le trait sélectionné soit supérieur à toutes les alternatives, à la plupart d'entre elles, ou seulement à quelques-unes

La théorie propensionniste de Bigelow et Pargetter (1987) rejoint la proposition de Ruse et semble apporter une réponse aux trois objections précédentes dans la mesure où ils interprètent le concept de fonction d'un trait ou d'une structure en termes de propension pour la sélection. Un trait ou une structure a une certaine fonction lorsque ses effets lui confèrent une propension à être sélectionné, sans que cette sélection ait besoin d'être effective. C'est-à-dire qu'un trait est doté d'une fonction dès sa première apparition[22], parce qu'il a une propension à la sélection, même si le hasard fait qu'en réalité il ne survive pas et n'ait aucun descendant. Mais avec un peu de chance, disent les auteurs, il *devrait* survivre, et s'il doit survivre c'est *parce que* il a une fonction. Puisqu'il n'est pas nécessaire, selon eux, d'avoir une histoire afin d'avoir une fonction, peu importe que les traits fonctionnels soient le fruit d'une évolution darwinienne ou d'une intervention divine. L'existence de fonctions biologiques ne dépend pas de la vérité de la théorie de l'évolution et il n'est pas nécessaire de connaître l'histoire d'un organe pour pouvoir lui attribuer une fonction.

Puisque la contribution à la *fitness* dépend de l'environnement, Bigelow et Pargetter accordent une importance particulière à la définition de l'habitat naturel d'une créature. Ce dernier, disent-ils (1987, p. 192), est celui dans lequel évoluent actuellement les créatures en question, à moins que l'environnement ne soit nouveau pour elles, auquel cas leur habitat naturel serait l'environnement précédent. L'une des particularités de la conception de ces auteurs est que la notion d'habitat n'est pas limitée à l'environnement des organismes entiers mais s'étend à l'environnement d'un organe dans l'organisme et à celui d'une cellule dans un organe.

Walsh (1996) fait une proposition similaire où la fonction d'un trait est sa contribution à la fitness individuelle moyenne dans un régime sélectif donné, lequel peut correspondre aussi bien à un environnement passé de ce trait qu'à un environnement plus récent ou actuel. Sa proposition vise à réunir dans une définition unique les conceptions historique et propensionniste, et peut-être aussi celle de Cummins (cf. Walsh & Ariew, 1996) :

---

d'entre elles. Il cite l'exemple d'un organisme qui se trouve être sélectionné alors que la plupart des alternatives étaient meilleures mais ont été accidentellement éliminées. L'interprétation propensionniste de la *fitness* a est due initialement à Brandon(1978) et Mills & Beatty (1979).

22 Walsh (1996, p. 563) formule une critique très pertinente à ce propos.

> The/a function of a token of type *X* with respect to selective regime *R* is to *m* iff *X*'s doing *m* positively (and significantly) contributes to the average fitness of individuals possessing *X* with respect to *R*. (Walsh, 1996, p. 564)

Cette définition relationnelle du concept de fonction prétend combiner les avantages des définitions historiques et anhistoriques et rendre possible l'attribution à un même trait des deux types de fonction sans contradiction aucune (1996, p. 570). C'est-à-dire qu'elle donnerait la possibilité d'appliquer à un même trait deux explications différentes : l'explication étiologique de la présence ou la prévalence d'un trait dans une population, et l'explication anhistorique de la *fitness* d'un individu ou d'un groupe en vertu de la contribution de ce trait à cette *fitness* (1996, p. 571).

Par ailleurs, l'attribution de la fonction à un type plutôt qu'à un *token* permet à la définition de conserver la dimension normative de l'approche étiologique, tandis que la référence à la *fitness* individuelle moyenne rend compte de la distinction entre conséquences fonctionnelles et accidentelles. Lorsque le régime sélectif fait référence à l'environnement actuel, la définition de Walsh coïncide avec celle de Bigelow et Pargetter, pour qui la contribution à la *fitness* doit être entendue de façon propensionniste. Lorsqu'il s'agit d'un environnement passé, la définition de Walsh rejoint celles de Millikan et de Neander dans la mesure où la sélection historiquement avérée d'un trait confirme sa contribution à la *fitness* des organismes porteurs. Mais contrairement aux définitions historiques précédentes, celle de Walsh autorise l'attribution d'une fonction dès la première génération, puisqu'elle dépend seulement de la contribution de ce trait à la *fitness* (au sens propensionniste) et non du résultat d'un long et hasardeux processus de sélection.

Pour illustrer les mérites de sa proposition, l'auteur donne les exemples suivants. *Zappus megauricularis imaginensis* (Fig. 3) est une souris imaginaire à grandes oreilles comportant deux groupes de population interfertiles ne s'étant pas croisés depuis de très nombreuses générations. La première population habite dans un milieu chaud, aride et sans prédateurs où les oreilles contribuent à réguler leur température. La seconde population vit dans un milieu montagneux plus frais où la taille et la forme des oreilles augmente leur acuité auditive et les aide à échapper aux prédateurs. Dans une première expérience de pensée (E1), un petit nombre d'individus est transporté des montagnes vers le désert où, après une brève période d'acclimatation, les nouveaux venus finissent pas trouver des aliments, des abris, des partenaires sexuels, et se comportent en tout comme leurs homologues autochtones. Dans la seconde expérience de pensée (E2), quelques individus sont pris dans les

deux populations et transportés ensemble dans un nouvel habitat plus humide où les souris n'ont pas à souffrir de la chaleur ni à craindre les prédateurs mais où il est difficile de trouver une nourriture appropriée. Cependant, le hasard fait que la forme des oreilles de *Zappus* ressemble aux plantes insectivores locales, de sorte qu'elles attirent de nombreuses mouches et deviennent ainsi une source abondante de nourriture. La question que pose alors Walsh est : « Quelle est la fonction des oreilles de *Zappus* dans chacun des exemples précédents ? »

Figure 3: Gerbille de Mongolie (*Meriones unguiculatus*) photographiée en milieu naturel dans le désert de Gobi. (Source : Morelle, 2007)

D'après l'auteur, la réponse des partisans d'une conception historique serait que les oreilles ont une fonction thermorégulatrice pour les souris du désert et une fonction auditive pour celles des montagnes (en admettant qu'il y ait eu « sélection pour » dans les deux cas). Dans le cas de E1, la réponse historique est insuffisante, dit Walsh, car les données biologiques actuelles étant identiques pour les souris autochtones et les nouvelles venues, et les oreilles apportant aux unes et aux autres les mêmes bénéfices, la conception étiologique leur attribue pourtant des fonctions différentes ; c'est-à-dire qu'elle attribue aux souris originaires des montagnes une fonction auditive qui dans le désert ne leur est d'aucune utilité. Dans le cas de E2, elle refuse de reconnaître aux oreilles une fonction qu'elle pourrait pourtant finir par leur attribuer après plusieurs générations. Et dans les deux cas, un biologiste appliquant une définition étiologique serait incapable d'identifier la fonction des oreilles d'une souris, à moins de connaître sa provenance, malgré sa connaissance par ailleurs exhaustive des données biologiques de l'animal : morphologie, physiologie, génome, etc.

La définition relationnelle, en revanche, permet d'attribuer aux oreilles de *Zappus* une fonction relative à son environnement d'origine, conformément à l'intuition historique, et une fonction relative à son environnement actuel, conformément à l'intuition propensionniste. Mais cette seconde interprétation de la *fitness*, comme nous l'avons indiqué, n'est pas partagée par tous les biologistes et philosophes de la biologie. De plus, le fait qu'elle dépende directement de la théorie darwinienne de l'évolution la rend inapplicable aux fonctions des artefacts.

La stratégie consistant à expliquer le concept de fonction en termes de *fitness* (dans son acception propensionniste) présente certains avantages mais aussi plusieurs inconvénients. Parmi ces derniers, on peut citer le flou et les controverses qui entourent ce terme, ainsi que la difficulté de son évaluation (M. Abrams, 2007; Ariew & Lewontin, 2004; Krimbas, 2004; Orzack & Forber, 2010; Rosenberg, 1983; Rosenberg & Bouchard, 2010; Sober, 2001). En effet, il est facile de dire *a posteriori* que l'apparition de tel ou tel trait au cours de l'évolution a augmenté les chances de survie et le succès reproductif des organismes qui en étaient pourvus, mais il est beaucoup plus difficile de prédire lequel parmi plusieurs individus similaires gagnera des batailles futures dont nous ignorons encore les conditions et les adversaires.

Une autre objection porte sur l'attribution de fonctions aux parties d'organismes qui, pour une raison ou pour une autre, ne peuvent pas se reproduire. Si les fonctions dépendent du succès reproductif, alors les individus stériles, par exemple, devraient en être privés. Doit-on en conclure que les organes des mules ne contribuent pas à sa *fitness* et que, dès lors, ils n'ont pas de fonction ? Cela est contre-intuitif. De manière analogue, imaginons qu'il ne reste plus que deux représentants vivants d'une espèce d'organismes sexués et que l'un des deux meure : que deviennent les fonctions des organes et des membres de l'individu restant puisqu'ils ne peuvent plus contribuer à sa *fitness*, car il n'a plus aucune chance de se reproduire ?

Nous disions plus haut que l'identification entre fonctions biologiques et adaptations ou contributions à la *fitness*, dans une perspective tournée vers le futur, semblait répondre aux trois objections formulées à l'encontre des théories étiologiques tournées vers le passé. Qu'en est-il ? La première objection s'appuyait sur l'exemple de William Harvey pour montrer qu'il est possible d'attribuer une fonction à un organe — ici le cœur — sans connaître son étiologie ni la théorie darwinienne. Les propositions que nous venons d'examiner semblent y apporter une réponse dans la mesure où elles ne font plus nécessairement dépendre les fonctions d'une histoire. Cela étant, dans la mesure où elles demeurent dépendantes de la théorie darwinienne, elles font référence aux concepts de *fitness* et de propension à la sélection. C'est-à-dire qu'elles ne rendent pas non plus compte du concept de fonction employé par Harvey.

La seconde objection disait que dans une conception fondée sur la sélection naturelle, il n'y aurait pas de fonctions biologiques si la théorie darwinienne s'avérait fausse. Les propositions précédentes présentent l'avantage de pouvoir attribuer les mêmes fonctions aux mêmes traits indépendamment du fait que leur origine soit naturelle ou surnaturelle, mais elles demeurent dépendantes des mécanismes invoqués par la théorie darwinienne. Si par exemple la survie et le succès reproductif étaient dictées par le hasard, alors les notions d'adaptation et de propension pour la sélection ou de contribution à la *fitness* perdraient de leur pertinence.

La troisième objection s'appuyait sur les doubles accidentels pour dire qu'il est contre-intuitif de refuser une fonction aux objets qui n'ont pas d'histoire (ou qui n'ont pas l'histoire qu'il faut). Là encore, les propositions précédentes semblent au premier abord y apporter une réponse dans la mesure où elles ne font pas dépendre les fonctions d'une histoire *passée*. Cependant, chez Ruse (1973a, p. 278), une adaptation implique non seulement que les organismes qui en sont dotés aient une plus grande disposition à survivre et à se reproduire (que des organismes autrement identiques), mais aussi que les organismes en question — passés, présents ou futurs — survivent et se reproduisent *effectivement* davantage grâce à ce trait. Autrement dit, il n'y a pas d'adaptation ni de fonction sans une réalisation historique de cette adaptation.

Par ailleurs, en mentionnant l'environnement relativement auquel se détermine la *fitness*, Bigelow et Pargetter donnent une épaisseur historique à leur définition. L'habitat naturel d'un organisme, disent-ils, est son environnement actuel, à moins que celui-ci ne soit nouveau, auquel cas ce serait l'environnement précédent. Mais combien de temps faut-il attendre pour qu'un nouvel environnement se transforme en l'habitat naturel d'un organisme ? Et quelles sont les causes de cette transformation ? Supposons qu'un trait *X*, sans incidence sur la survie et la reproduction d'un organisme dans son habitat naturel, contribue à sa *fitness* quand on le transplante dans un autre environnement. Sachant que cet environnement deviendra sans doute le nouvel habitat naturel des organismes transplantés, le trait *X* a-t-il une fonction ? Si oui, a-t-il une fonction en vertu de l'histoire future qu'on lui prête ?[23]

Chez Ruse (1973a), peut-être aussi chez Griffiths (1993) et Buller (1998), et de manière générale chez tous ceux qui associent la fonction d'un trait à un moment t_1 au succès *avéré* de ce trait ou des organismes porteurs à un moment t_2, les attributions fonctionnelles correspondent à ce que Arthur Danto, dans sa *Philosophie analytique de l'histoire* (1965), appelle des phrases narratives[24]. Leur caractéristique principale est de faire référence à deux événements séparés dans le temps

23 Godfrey-Smith formule une critique similaire portant sur les rivaux : Sachant que la *fitness* est une notion comparative, les rivaux par rapport auxquels on peut mesurer le succès reproductif d'un organisme sont forcément des organismes présents ou passés, c'est-à-dire que la contribution présente et future d'un trait à la *fitness* de l'organisme porteur ne peut se faire que par rapport à des traits alternatifs présents ou passés : « *The range of alternatives the trait has a propensity to be selected over are the ones it actually triumphed over, and continues to be selected over.* » (1994, p. 352-3)

24 Dans une précédente publication (Molina Pérez, 2006, p. 53-55), la portée de cet argument était beaucoup plus large. Après une réflexion plus approfondie, il nous a semblé que sa validité n'était pas aussi générale que nous le pensions.

tout en ne portant que sur le plus ancien des deux. Par exemple, la phrase « L'auteur de *Don Quichotte* est né en 1547 » n'avait aucun sens en 1547 ni en 1560 ; et elle n'avait pas le même sens en 1605, année de publication du roman, qu'en 2016. Peu à peu, la naissance de Cervantes en 1547 a acquis, dans le cadre d'une structure temporelle plus vaste, une signification historique qu'aucun chroniqueur de l'époque ne pouvait connaître. Même s'il existait ce que Danto appelle un Témoin Idéal, c'est-à-dire un chroniqueur omniscient capable d'écrire en temps réel dans un Journal tout ce qui se passe dans le monde, sa description des faits serait nécessairement incomplète, car elle ne tiendrait pas compte de leur signification. Cette signification est relative à une histoire, laquelle est décrite de manière narrative :

> « To ask for the significance of an event, in the *historical* sense of the term, is to ask a question which can be answered only in the context of a *story*. The identical event will have a different significance in accordance with what set of *later* events it may be connected » (Danto, 1965, p. 11)

Comme la signification d'un fait historique, la fonction d'un trait dans la conception de Ruse est définie à la lumière de ses effets au cours des générations postérieures à sa première apparition. On peut donc dire : « Le trait a acquis la fonction f à t_1 » bien qu'à ce moment-là il ne possédait pas encore la fonction qu'on lui attribue, de même que Miguel de Cervantes n'était pas encore l'auteur de *Don Quichotte* à sa naissance en 1547. C'est seulement depuis une perspective plus récente, avec un certain recul, que nous pouvons lui attribuer cette signification. Il n'y a en réalité aucun moment précis où le trait ait acquis la fonction qu'on lui attribue, mais on peut désigner un moment t_2 où le trait a déjà acquis cette fonction (Fig. 4). Quand on dit alors qu'un cœur mal formé, à t_3, a la fonction de pomper le sang, on le considère depuis la perspective d'une structure historique dont le début se situe entre t_1 et t_2. Cette structure est l'étiologie de l'organe, et c'est elle qui détermine sa fonction à t_3.

Il ne fait aucun doute que les faits que les historiens rapportent ont réellement eu lieu dans le passé. Et il ne fait aucun doute qu'un démon de Laplace ou un chroniqueur idéal pourrait en donner une description exhaustive ; mais cette description placerait tous les faits sur le même plan, de telle sorte que la découverte de l'Amérique par Christophe Colomb n'aurait pas plus d'importance à ses yeux que la chute d'une pomme. Ce qui transforme un fait spatio-temporel en événement historique est sa signification. De même, il ne fait aucun doute que les fonctions ont une dimension spatio-temporelle, mais elles ne sont pas elles-mêmes des réalités spatio-temporelles ou physico-chimiques. Elles pourraient être essentiellement de l'ordre de la signification.

Pour finir, nous retiendrons trois choses qui, de notre point de vue, sont les points forts des conceptions propensionnistes vis-à-vis des définitions étiologiques concurrentes :

- la prise en compte de la contribution d'un trait à l'organisme qui le possède ;

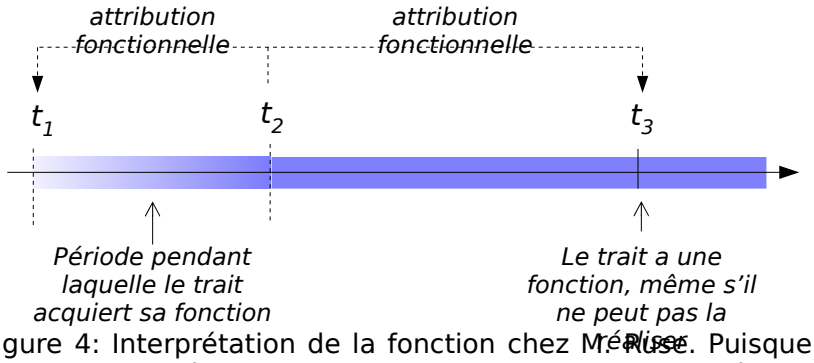

Figure 4: Interprétation de la fonction chez M. Ruse. Puisque la fonctionnalité d'un trait lors de son apparition (t_1) dépend de son succès futur, on ne peut l'attribuer que depuis le moment où celui-ci est avéré (t_2). Les attributions fonctionnelles postérieures (t_3) sont relatives à l'acquisition de la fonction pendant la période comprise entre t_1 et t_2.

- la prise en compte de l'environnement[25] ; et
- l'indépendance relative de l'attribution fonctionnelle à l'égard de l'histoire.

Mais la principale vertu de ces propositions est sans doute la mise en évidence d'une lacune explicative des approches historiques : en décrivant la fonction d'un trait comme étant l'effet pour lequel il a été sélectionné, elles n'expliquent pas pourquoi il a été sélectionné ; elles ne rendent pas compte de la valeur de ce trait pour l'organisme qui le porte.

4. La place des valeurs dans l'explication fonctionnelle

4.1. Deux fonctions, trois explications

Les attributions fonctionnelles permettent de répondre à des questions de type « pourquoi ? » relatives au comportement ou à la présence d'un item dans un système. Dans le domaine de l'éthologie, un même comportement peut s'expliquer en termes de :

1. *mécanismes* physiologiques et de stimuli physiques responsables du comportement ;
2. *valeur de survie* (ou reproductive) actuelle du comportement ;
3. *histoire* évolutive du comportement ;
4. *développement* du comportement au cours de la vie de l'individu[26].

Dans le cadre d'une approche étiologique tournée vers le passé, attribuer une fonction à un comportement ou à un trait c'est expliquer sa présence à travers son histoire évolutive (3), tandis que dans une approche tournée vers le présent ou vers le futur, l'explication porte plutôt sur sa valeur de survie ou sa valeur reproductive (2). C'est à ces deux questions différentes que la proposition de Walsh (1996) cherche à répondre simultanément. Et c'est également pour différencier ces deux questions que Godfrey-Smith (1994) distingue l'histoire récente d'un trait de son histoire plus lointaine : on explique ainsi que les manchots aient des ailes parce qu'ils les ont héritées de leurs ancêtres non-aqua-

25 Il convient néanmoins de signaler la critique que Alvin Plantinga (1993, p. 206) formule à propos du recours à l'environnement. D'après lui, la notion d'habitat naturel à laquelle recourent Bigelow et Pargetter dans leur analyse du concept de fonction implique elle-même déjà ce concept, de sorte que leur analyse s'avère être circulaire.

26 Cité par Godfrey-Smith (1994, p. 351), repris et commenté par Walsh (1996, p. 557) et attribué à Timbergen.

tiques capables de voler (2), et parce qu'elles leur permettent actuellement de se propulser dans l'eau et hors de l'eau (3).

Considérons les questions suivantes où chaque réponse donne lieu à une nouvelle interrogation :

Q1 : Pourquoi certaines plantes sont-elles vertes ?

R1 : Parce qu'elles contiennent de la chlorophylle qui est de couleur verte, laquelle s'explique à son tour par la structure physique et les propriétés optiques de cette molécule.

Mais cette réponse ne fait que déplacer le problème que l'on peut reformuler de la manière suivante :

Q1' : Pourquoi certaines plantes contiennent-elles de la chlorophylle ?

R1' : Parce que la production de cette molécule fait partie du programme génétique de ce type de plantes. Une réponse que l'on pourrait compléter en détaillant les mécanismes causaux qui en sont responsables à l'échelle moléculaire.

Mais cette réponse, aussi précise et détaillée soit-elle, reste insuffisante. Pour aller plus loin, il faudrait donner une explication de l'existence des gènes et des mécanismes moléculaires impliqués ; par exemple, une explication historique disant que les plantes actuelles contiennent de la chlorophylle parce qu'elles ont hérité cette caractéristique de leurs ancêtres.

Est-ce suffisant ? Non, car cette réponse ne fait que repousser le problème dans le temps. Daniel Dennett (1995, p. 102-103) prend l'exemple du cou des girafes pour arriver à la même conclusion. Pourquoi les girafes ont-elles un long cou ? Parce qu'elles l'on hérité de leurs parents, qui l'on hérité de leurs parents, et ainsi de suite pendant des millions d'années… jusqu'à aboutir à de très lointains ancêtres qui n'avaient même pas de cou. Fin de l'explication. Et si vous n'êtes pas satisfait, ajoute Dennett, vous le seriez encore moins si la réponse rentrait dans tous les détails de l'histoire développementale et nutritionnelle de chacune des girafes de la lignée. Cela nous conduit à formuler deux nouvelles questions :

Q2a : Pourquoi les ancêtres des plantes vertes actuelles contenaient-ils de la chlorophylle ?

R2a : Il y a au moins deux réponses possibles à cette question. Si l'on fait référence à la première génération d'organismes porteurs de ce trait, la réponse est sans doute : par hasard. Si l'on fait référence aux générations postérieures, la réponse est la même que pour les plantes actuelles : parce qu'elles ont hérité cette caractéristique de leurs ancêtres.

Q2b : Pourquoi les plantes vertes actuelles ont-elles hérité cette caractéristique de leurs ancêtres ?

R2b : Dans le cadre de la théorie darwinienne, on devrait dire que ce trait est le produit d'un processus de sélection naturelle où les individus le possédant ont bénéficié d'un taux de survie et de reproduction suffisant pour qu'il soit transmis de génération en génération jusqu'à aujourd'hui.

Est-ce suffisant ? Non, car outre le fait avéré de sa transmission, nous ne savons pas si ce trait a contribué au succès reproductif des organismes porteurs ni comment il y a éventuellement contribué. En effet, on pourrait donner la même réponse pour le bruit des battements de cœur : les cœurs actuels font du bruit à cause du succès reproductif de nos ancêtres. Et cette réponse resterait insuffisante même si l'on connaissait dans le détail l'histoire (la vie, la mort, le nombre de descendants, etc.) de tous les ancêtres individuels des plantes vertes actuelles, car le fait qu'un trait ait été sélectionné mérite lui-même une explication et suscite une nouvelle question :

Q3 : Pourquoi la chlorophylle a-t-elle été sélectionnée ?

R3 : Parce qu'elle permet de réaliser la photosynthèse, laquelle conférait aux organismes porteurs un avantage sélectif que l'on pourrait exprimer en termes de valeur adaptative ou de contribution à la *fitness*. En d'autres termes, parce que cette capacité améliorait leurs chances de survie et de reproduction[27].

Cette dernière explication fait référence à l'utilité de la chlorophylle par le passé et elle reste valable pour les plantes actuelles, de sorte qu'elle permet de répondre à la première question (Q1). D'autant plus que la photosynthèse en C4 est apparue indépendamment plus de 60 fois au cours de l'histoire. Il faudrait donc expliquer, non pas seulement pourquoi elle est apparue et a été sélectionnée une fois, mais pourquoi autant de fois de façon indépendante. Pour être complète, il faudrait y ajouter une description des bénéfices comparatifs de la photosynthèse.

Les explications mécanistes, qui invoquent des causes proximales (R1) et (R1'), sont toujours vraies, mais elles ne sont pas suffisantes, car elles ne rendent pas compte de l'existence de ces mécanismes causaux eux-mêmes. Sont-ils le fruit du hasard, d'un dessein ou de quelque chose d'autre ? Les explications historiques, qui invoquent des causes distales (R2a) et (R2b), sont également vraies pour les organismes biologiques.

27 Avec la réponse précédente, on pouvait se limiter à dire que les individus dotés de chlorophylle (et donc capables de réaliser la photosynthèse) ont *de fait* eu un succès reproductif comparativement supérieur. Ici on s'interroge sur la raison de ce succès, en supposant qu'il n'est pas purement et simplement aléatoire, et on indique que la chlorophylle (c'est-à-dire la photosynthèse) en est responsable.

Elles expliquent l'existence des mécanismes causaux actuels comme étant le fruit d'une sélection naturelle entre différentes alternatives. Mais elles ne sont pas suffisantes, car elles n'expliquent pas pourquoi une alternative donnée a été sélectionnée au détriment des autres[28]. Or, c'est précisément ce que cherchent à faire les explications adaptationnistes et propensionnistes, montrer que si tel trait a été sélectionné parmi d'autres ce n'est pas par hasard mais parce qu'il conférait à son porteur un avantage sélectif[29].

En généralisant les réponses précédentes, on peut distinguer au moins trois niveaux hiérarchiques d'explication concernant l'existence d'un trait organique (organe, structure, processus). Et ces trois niveaux d'explication étaient implicitement présents dans la conception étiologique de Wright :

(a) l'explication mécanique (en termes de causes proximales) ;

(b) l'explication historique (en termes de causes distales) ;

(c) l'explication valorative (en termes de « valeurs »).

Chacun de ces trois niveaux rend compte du précédent : l'existence actuelle d'un trait s'explique par des mécanismes causaux proches (a), lesquels s'expliquent par des mécanismes causaux passés (b), lesquels s'expliquent à leur tour par la « valeur » du trait en question (c). Ils restent néanmoins indépendants, car on peut donner une explication à un certain niveau sans connaître celle des niveaux inférieurs et supérieurs. Ainsi, une explication valorative de l'existence d'un trait (c) resterait vraie même si son explication historique (b) ne l'était pas.

Par exemple, si l'œil n'était pas le produit de la sélection naturelle mais d'un dessein intelligent, les raisons pour lesquelles il a été intentionnellement conçu seraient sans doute analogues aux raisons pour lesquelles nous croyons qu'il a été naturellement sélectionné au cours de l'évolution darwinienne. De plus, étant donné qu'il n'y a pas un œil mais plusieurs, lesquels sont apparus et ont évolué indépendamment les uns des autres, l'explication valorative de l'existence d'un organe de type « œil » est indépendante des explications historiques des différents types d'yeux. On

28 Lorsqu'un trait est favorisé par la sélection naturelle et se diffuse à l'ensemble de la population, il est probable qu'il soit adaptatif et qu'il accroisse le succès reproductif des organismes qui le portent, mais les raisons pour lesquelles ce trait est adaptatif et contribue à la *fitness* n'ont pas besoin d'être connues et pourraient éventuellement être dictées par le hasard.

29 Dans une perspective propensionniste, on peut dire que chaque trait, à un moment donné et dans un environnement donné, contribue plus ou moins à la *fitness* de l'organisme qui le porte, et on en déduit qu'il a plus ou moins de chances d'être sélectionné au cours du temps. Dans une perspective historique, on constate la sélection d'un trait au cours du temps, et on en déduit qu'il a contribué plus ou moins à la *fitness* des organismes porteurs. L'ordre de l'explication est inverse.

pourrait en dire de même de la photosynthèse en C4 qui est apparue indépendamment plus de 60 fois.

4.2. Explications causales et non causales

On pourrait exprimer ces raisons, de façon générale, en termes d'utilité (Ayala, 1970; Canfield, 1964). Et c'est cette utilité manifeste de l'organe qui peut nous pousser à lui attribuer une fonction comme nous le faisons pour les artefacts[30] : si les protubérances latérales de pseudo-lions issus d'un accident cosmique leur servent principalement à marcher, courir et saisir des proies de manière efficace, et si cela contribue à leurs chances de survie et de reproduction, alors il n'y a *a priori* aucune raison de leur attribuer une fonction différente de celle-là. De même, la fonction qu'on attribue aux membres latéraux des manchots est de leur permettre de se déplacer à la fois sur terre et dans l'eau. Quand on découvre que les manchots sont des oiseaux issus d'ancêtres communs aux albatros actuels, on peut attribuer à ces membres une autre fonction qu'ils ne sont plus capables de réaliser aujourd'hui, celle de voler. Cette seconde attribution fonctionnelle donne à la fois une explication historique de l'existence de ces membres chez les manchots actuels (b), et une explication « valorative » des membres correspondants chez leurs ancêtres aériens (c).

Bien sûr, le fait d'être utile, d'une manière ou d'une autre, ne suffit pas à expliquer l'existence d'un trait organique ni celle d'un item manufacturé. Il leur faut nécessairement une cause, laquelle peut être aussi bien l'action d'un créateur intelligent qu'un processus non intentionnel comme la sélection naturelle. Dans les deux cas, l'existence actuelle d'un

30 Quand on s'interroge sur l'existence et les caractéristiques d'un artefact ou de l'une de ses parties, la fonction qu'on lui attribue au premier abord correspond habituellement à la valeur pratique ou esthétique qu'on lui reconnaît, car c'est généralement cette valeur qui explique son existence. Chez les êtres vivants, l'utilité d'un trait pour l'organisme est également, dans la plupart des cas, ce qui explique son existence. Francisco Ayala (1970) propose d'employer le critère d'utilité pour distinguer les entités téléologiques de celles qui ne le sont pas, lesquelles incluent aussi bien les organismes biologiques que les objets manufacturés : « *A feature of a system will be teleological in the sense of internal teleology if the feature has utility for the system in which it exists and if such utility explains the presence of the feature in the systems. Utility in living organisms is defined in reference to survival or reproduction. A structure or process of an organism is teleological if it contributes to the reproductive efficiency of the organism itself, and if such contribution accounts for the existence of the structure or process. Man-made tools or mechanisms are teleological with external teleology if they have utility, i.e., if they have been designed to serve a specified purpose, which therefore explains their existence and properties. If the criterion of utility cannot be applied, a system is not teleological.* » (1970, p. 12-13)

trait ou d'un item ne signifie pas nécessairement que celui-ci soit plus utile ni plus adaptatif ni meilleur d'une manière ou d'une autre que ceux qui ont disparu. De nombreux facteurs accidentels peuvent contribuer à ce que, parmi plusieurs traits concurrents, les plus adaptatifs disparaissent et les autres soient sélectionnés. L'existence actuelle d'un trait ou d'un item a en effet toujours une cause, mais pas toujours une raison — c'est-à-dire une justification rationnelle, — et le fait de savoir comment ce trait est arrivé jusque-là ne permet pas forcément de comprendre pourquoi. En revanche, lorsque le succès a une raison d'être, on peut en comprendre le pourquoi sans savoir le comment. Par conséquent, si l'utilité d'un trait ou d'un item n'est pas suffisante pour expliquer son existence, la description des causes ne l'est pas davantage.

Dans la plupart des cas, l'histoire causale particulière d'un trait organique nous est totalement inconnue et elle s'avère inutile pour déterminer sa fonction. C'est la raison pour laquelle les conceptions étiologiques tournées vers le passé font référence au mécanisme général de la sélection naturelle pour justifier les attributions fonctionnelles : les effets d'un trait sont fonctionnels *si et seulement si* il y a eu sélection naturelle de ce trait pour ses effets, indépendamment de la manière dont s'est déroulée la sélection. De sorte qu'une diversité infinie d'histoires sélectives sont compatibles avec la même fonction. Cela reste une explication causale dans la mesure où la sélection naturelle est un mécanisme causal, mais elle se situe à un niveau de généralisation supérieur.

L'explication alternative dont nous essayons ici de dessiner les contours se situe à un niveau encore plus élevé, car elle fait abstraction du mécanisme causal effectivement responsable de l'existence de l'item ou du trait organique en question. C'est ce qui caractérise les explications fonctionnelles, comme l'indiquait déjà Larry Wright :

> « [E]stablishing one [functional explanation] merely indicates the presence of a selection background of some kind and it leaves an enormous amount of theoretical detail completely open. The exact physical mechanism, the precise details of the selection process, in principle even the type of selection (natural or conscious), are not determined merely by establishing a functional ascription-explanation. » (L. Wright, 1976, p. 108-9)

L'un des avantages de ce type d'explication par rapport à une explication causale plus précise réside justement dans le fait qu'elle peut être formulée alors que les causes ne nous sont pas encore connues et qu'elle fournit, de ce fait, de précieuses informations pour l'investigation ultérieure d'un phénomène. Le problème principal des conceptions historiques de la téléologie biologique réside peut-être dans le fait qu'elles attachent trop d'importance au mécanisme causal qui en est responsable et perdent de vue la valeur heuristique du langage fonctionnel :

> « Any analysis that leans too heavily on details of current evolutionary theory, for instance, is bound to misrepresent one of the most important aspects of function-talk. [Attributions of function are better explanations for the origin of the functional trait than an evolutionary narrative which contains far more information.] Of course there *are* purposes for which the underlying account is better than the functional one. But this is just what we should expect: they are not rivals. They do different jobs. » (L. Wright, 1976, p. 109-10)

Chez Wright, l'idée de sélection implicitement présente derrière le premier critère (*X is there because it does Z*) visait à distinguer les effets fonctionnels et accidentels de la présence d'un item X — en particulier ses effets accidentellement utiles (contre Canfield) — et à généraliser la définition aux fonctions biologiques et techniques, mais elle n'avait pas pour vocation de se transformer en pierre de touche de l'approche étiologique. Lorsqu'il souligne que la sélection dont il est question est une sélection en vertu d'un avantage résultant, il laisse entendre que ce n'est pas seulement le fait d'avoir été sélectionné qui compte pour attribuer une fonction (naturelle ou artificielle), mais aussi et surtout la raison sous-jacente à cette sélection, c'est-à-dire l'avantage qui en résulte :

> « Let me refer to selection by virtue of resultant advantage of this sort as "consequence-selection." Plainly, it is this kind of selection, as opposed to mere discrimination, that lies behind conscious functions: the consequence *is* the function. Equally plainly, it is specifically this kind of selection of which *natural* selection represents an extension. » (L. Wright, 1973, p. 163)

C'est cet avantage résultant qui explique pourquoi un item a été sélectionné, et donc pourquoi il existe. Faire exclusivement référence au mécanisme de la sélection naturelle sans mentionner l'avantage qui la justifie serait donner une explication causale — et pas une explication téléologique — de la présence de l'item en question. Certes, la sélection naturelle opère sur la base de ce type d'avantages, de sorte que sa mention dans les définitions historiques peut être interprétée comme une référence plus ou moins indirecte à un avantage résultant, mais les propositions de Millikan et de Neander se concentrent sur le *processus* de sélection, pas sur l'avantage ni sur le résultat.

4.3. La loterie à Babylone

Mark Bedau (1991) a inventé un contre-exemple à l'approche étiologique destiné à réfuter les tentatives de naturalisation de la téléologie qui s'appuient exclusivement sur la sélection naturelle. L'auteur imagine une planète inhabitée où des cristaux d'argile sont soumis à un processus évolutif par sélection darwinienne. Bien que certaines des caractéristiques de ces cristaux aient été sélectionnées en vertu de leurs effets, de la même

manière que les traits des organismes biologiques, le discours téléologique ne s'applique pas dans leur cas. L'argument consiste à dire que contrairement à ce que prétendent les conceptions étiologiques basées sur la sélection naturelle, celle-ci n'est pas suffisante pour rendre compte de la téléologie.

Les raisons pour lesquelles la téléologie ne s'applique pas aux cristaux d'argile ne sont pas très claires. L'auteur fait d'abord appel à l'intuition pour dire que les caractéristiques des cristaux ne sont pas le genre de choses qui appartiennent au domaine de la téléologie. Ensuite, il se pose la question : « qu'est-ce qui différencie la sélection naturelle chez les cristaux et chez les organismes biologiques, et pourquoi la téléologie est-elle présente chez les uns et pas chez les autres ? » La réponse de l'auteur est que la téléologie implique des valeurs et que celles-ci ne s'appliquent pas aux objets inertes : on peut dire que quelque chose est bon ou mauvais pour un cheval, un oiseau, une plante ou une amibe, mais pas pour un caillou pour ni un cristal d'argile. Ce qui n'est pas clair dans son explication, c'est la raison pour laquelle les notions de valeur s'appliquent à certains objets (les êtres vivants) et pas à d'autres (la matière inerte).

Pour palier cet inconvénient et compléter l'argument de Bedau sans en tirer nécessairement les mêmes conclusions, il faudrait imaginer une situation où la sélection naturelle agirait sur les êtres vivants et où, cependant, on pourrait difficilement parler de téléologie ou de fonctions à leur propos. Une nouvelle de Jorge Luis Borges, intitulée « La loterie à Babylone » et tirée de son recueil *Ficciones* (1944), nous en offre la trame[31].

31 Borges imagine une Babylone où la loterie s'est tellement développée qu'elle y a acquis un statut presque métaphysique. Elle est secrète, gratuite et générale. Tous les personnes libres participent automatiquement à des tirages sacrés tous les soixante jours qui déterminent leur destin jusqu'au tirage suivant. Un bon numéro peut rapporter au gagnant de se voir élu au conseil des mages ou l'emprisonnement d'un ennemi. Un mauvais numéro, en revanche, peut le condamner à la mutilation, à l'infamie ou à la mort. Les récompenses et les sanctions possibles sont innombrables. N'importe quel fait de la vie quotidienne peut être le résultat d'un tirage de la loterie. Certains sont le résultat de plus de trente ou quarante tirages. Les combinaisons de ce genre sont ardues ; mais les membres de la Compagnie (c'est ainsi que s'appelle l'institution qui organise la loterie) sont tout-puissants et rusés. Des erreurs ne manquent pas de se produire dans l'attribution des prix, mais elles corroborent le hasard qui est la base de la loterie. L'exécution des résultats de la loterie, jusqu'à ses aspects les plus infimes, est elle-même sujette au hasard sous la forme de nouveaux tirages dont la complexité tend à l'infini. Il existe aussi des tirages impersonnels, à finalité indéfinie, qui dictent qu'un oiseau soit lâché du haut d'une tour ou que chaque siècle un grain de sable soit prélevé (ou ajouté) parmi les innombrables grains qu'il y a sur la plage. Les résultats de la loterie étant secrets, nul ne sait si un événement quelconque, du plus simple au plus terrible, est un résultat de la loterie. À vrai dire, nul

Imaginons qu'à Babylone, dès leur naissance, tous les individus participent à une Loterie dont les « prix » peuvent être aussi bien des récompenses que des sanctions. Un bon numéro peut rapporter aux gagnants de trouver un meilleur emploi, de gagner une élection politique ou d'épouser la personne de leur choix. Un mauvais numéro, au contraire, peut valoir une peine de prison, l'inoculation d'une maladie, et même une condamnation à mort. La liste des « prix » pour chaque tirage est indéfiniment longue et extrêmement variée. Chaque individu dispose d'un unique billet de Loterie comportant plusieurs milliers de numéros qu'il n'a pas choisis ni ne peut changer et avec lesquels il participe à tous les tirages. Parfois, un seul des numéros de son billet suffit pour recevoir un prix, parfois il en faut des dizaines. Le nombre de prix en jeu est tellement grand qu'un même individu en reçoit toujours plusieurs à chaque tirage, certains sont bons, d'autres mauvais, d'autres relatifs aux circonstances (les manteaux de fourrure, par exemple, sont utiles aux habitants des zones les plus froides). Beaucoup estiment que la distribution des prix n'est pas équitable, c'est-à-dire pas tout à fait aléatoire. Certains murmurent même que les dés seraient pipés. Les réductions d'impôts, par exemple, retombent souvent sur le numéro 6 tandis que l'amputation des deux jambes retombe périodiquement sur une même combinaison de trois numéros : 8, 248 et 972. Peut-on affirmer pour autant que le 6 sert à réduire les impôts des individus qui le possèdent, ou qu'il existe pour cela ? Peut-on lui attribuer cette fonction ? À ce stade, sans doute pas.

Chaque individu hérite de ses parents, à parts égales, les numéros qui composent son Billet de Loterie Personnel. Cette transmission est également régulée par des tirages au sort qui déterminent quels nombres seront transmis à l'enfant. Les erreurs ne sont pas rares : confusions, répétitions, mauvaises transcriptions, etc. Il arrive qu'un individu ait dans son billet un numéro n'appartenant à aucun de ses parents, parfois même un numéro totalement nouveau, jamais tiré auparavant. La rumeur prétend que ces numéros-là n'apportent rien de bon. De fait, beaucoup disparaissent aussi vite qu'ils sont apparus. Mais les Babyloniens n'aiment pas les spéculations ni les calculs de probabilités et la plupart n'associent pas le fait d'avoir tels ou tels numéros aux chances de trouver un emploi ou de finir en prison. Ils s'abandonnent au hasard. Pourtant, certains groupes de numéros finissent par être plus répandus que d'autres. Il est plus facile de trouver des individus ayant le 6 que le 179. Certains numéros sont d'ailleurs tellement répandus que la quasi-totalité des Babyloniens les ont dans leur billet. D'autres sont partagés par les habitants d'un même quartier, ou par les membres d'une même profession. En général il ne s'en plaignent pas. Les habitants du Nord gelé, par exemple, reçoivent davantage de manteaux que leurs voisins du Sud

---

ne sait si la Compagnie existe toujours, ni même si elle a jamais existé. Et peu importe, car Babylone n'est pas autre chose qu'un infini jeu de hasard.

torride. Peut-être parce que ceux qui n'en reçoivent jamais finissent par déménager ou par mourir de froid.

L'étude des Chroniques Anciennes révèle que la distribution des numéros et des combinaisons de numéros parmi la population n'a pas toujours été la même. Certains ont augmenté leur présence, d'autres ont disparu. Le 179 par exemple se fait rare, ainsi que la combinaison 8, 248, 972. Les Chroniques montrent également que les individus dont le billet comporte un 6 n'ont pas toujours été disposés à gagner davantage d'exemptions d'impôts que les autres. Il y a très très longtemps, le 6 avait plutôt tendance à rapporter un titre de noblesse ou une charge publique. Il en va de même pour beaucoup d'autres nombres et combinaisons. Des experts étrangers, Hindous pour la plupart, ont étudié la Loterie de Babylone et proposé plusieurs explications aux apparentes irrégularités du hasard. Certains affirment que le temps est infini, que la Loterie existe depuis toujours et que les différences de répartition des prix et des numéros sont dues à notre myopie historique ; si nous étions capables d'embrasser du regard les siècles des siècles, nous verrions que chaque prix est attribué le même nombre de fois à chaque numéro, et que tous les numéros rapportent exactement les mêmes prix. D'autres prétendent que les dés sont pipés à cause de l'imperfection des machines, qui favorisent certains numéros pour certains prix, et que l'usure ou les réparations successives font évoluer ces répartitions. D'autres finalement, expliquent que les tirages ordinaires sont effectivement biaisés, mais que ces biais sont déterminés par des tirages extraordinaires dont les biais sont déterminés par des méta-tirages extraordinaires dont les biais sont déterminés par d'autres méta-tirages… et ainsi de suite jusqu'à la fin.

La présence du numéro 6 dans un billet de Loterie a notamment pour conséquence de réduire les impôts des individus qui le possèdent, ou plutôt de les réduire davantage que les autres numéros. Les réductions d'impôts contribuent à la richesse et au succès reproductif de ceux qui en bénéficient, et on peut observer comment, au fil des générations et à cause des effets de sa présence dans les billets de Loterie, le 6 s'est diffusé rapidement au sein de la population au détriment d'autres numéros. Actuellement, on peut affirmer sans crainte de se tromper que les individus dont le billet de Loterie comporte un 6 ont statistiquement plus de chances que le reste de la population de réussir dans la vie et d'avoir une descendance nombreuse. On peut même aller jusqu'à dire que le 6[32] a historiquement été sélectionné (parmi d'autres) à cause des effets statistiquement favorables de sa présence dans les billets, et, qu'à l'heure actuelle, les effets de la présence de ce numéro dans les billet augmentent sa propension à la sélection. Peut-on affirmer pour autant que le 6 sert à réduire les impôts des individus qui le possèdent, ou qu'il existe pour

32 Ou alors les individus dont les billets de Loterie comportent un 6.

cela ? Peut-on lui attribuer cette fonction ? Intuitivement, certainement pas.

Hérédité, variation, sélection, adaptation. La Loterie à Babylone remplit les conditions de base de la sélection naturelle et s'ajuste aux définitions étiologiques basées dessus, qu'elles soient tournées vers le passé, le présent ou le futur. Pourtant, les numéros des billets de loterie ne sont pas le genre de choses qui appartiennent au domaine de la téléologie, comme dirait Bedau. L'intuition se refuse à leur attribuer une fonction. Pourquoi ? La réponse de cet auteur ne nous est d'aucune aide car la dimension valorative est bel et bien présente dans le cas de la Loterie à Babylone : les effets de la présence d'un numéro dans les billets de Loterie sont bons ou mauvais pour les individus auxquels ils appartiennent, de même que les effets de la présence d'un gène dans le génome d'un organisme quelconque. Pourquoi le discours téléologique est-il applicable dans un cas et pas dans l'autre ?

Chapitre II

Approche systémique (synchronique)

Face à l'approche étiologique, on trouve un ensemble de théories que l'on peut rassembler sous le nom d'approche systémique. Elles ont pour caractéristique commune d'analyser la téléologie et les fonctions biologiques du point de vue la théorie des systèmes. Et elles s'intéressent moins à la signification des concepts qu'à l'explication naturaliste du fonctionnement des systèmes auxquels on attribue des fonctions ou dont le comportement peut être décrit en termes téléologiques.

Hormis les partisans — biologistes pour la plupart — d'une stratégie centrée sur la notion de programme (génétique), les défenseurs de l'approche systémique cherchent à éviter toute distinction entre les organismes biologiques et les autres systèmes physiques. Peut-être, comme dit Ernst Mayr (1992, p. 120), par peur d'ouvrir la porte à des considérations d'ordre métaphysique ou non-empiriques.

L'approche systémique s'inspire en premier lieu de la cybernétique. Elle a notamment été défendue par Ernest Nagel (1961, 1977a, 1977b) dans ce qui demeure l'une des références incontournables des stratégies de naturalisation de la téléologie biologique. Nagel cherchait à rendre compte, de façon générale, des systèmes dont le comportement semble dirigé vers un but (*goal directed behaviour*). Elle a aussi été défendue de manière consistante par Christopher Boorse (1976, 2002).

L'approche systémique a ensuite été défendue, contre l'approche étiologique, par Robert Cummins (1975) sous la forme d'une théorie dispositionnelle. La conception du rôle causal (*causal role conception*) de Cummins a été reprise par de nombreux auteurs et est devenue la principale alternative à la conception des effets sélectionnés (*selected effects conception*) de Larry Wright. Parmi les héritiers de cette approche, nous

accorderons une attention particulière à la proposition de Paul Sheldon Davies (2001).

1. Finalité comme programme

Plusieurs biologistes de renom (Jacob, 1970; Maynard Smith, 2000; Mayr, 1961, 1974, 1992; Monod, 1970; Monod & Jacob, 1961) ont proposé de comprendre la téléologie biologique à partir de notions comme celle de *programme* génétique[33] :

> Longtemps le biologiste s'est trouvé devant la téléologie comme auprès d'une femme dont il ne peut se passer, mais en compagnie de qui il ne veut pas être vu en public. À cette liaison cachée, le concept de programme donne maintenant un statut légal.
> (Jacob, 1970, p. 17).

Cette interprétation présente l'avantage de rendre compte de la finalité apparente des processus biologiques en termes de mécanismes physicochimiques tout en préservant l'espace et la légitimité d'un discours dit « téléonomique » propre à la biologie.

Le concept de téléonomie, formulé pour la première fois par Colin S. Pittendrigh en 1958, a été repris avec des significations différentes par plusieurs biologistes et philosophes de la biologie pour éliminer les connotations inacceptables du concept de téléologie. Ernst Mayr le définit comme suit : « *a teleonomic process or behavior is one that owes its goal-directedness to the operation of a program* ». Le même auteur définit la notion de programme comme étant « *coded or prearranged information that controls a process (or behavior) leading it toward a given end* » (1974, 1992, p. 127). De manière plus générale, la téléonomie est la reconnaissance d'une finalité *apparente* dans les phénomènes naturels en dehors de toute doctrine ou présupposition métaphysique. C'est une sorte de fiction méthodologique rendant possible — et surtout légitime — la recherche d'une explication scientifique.

Henri Atlan (1979, p. 16) disait que cette interprétation de la finalité situe les explications biologiques dans le cadre encore mal défini d'une physicochimie non classique qui mêle la thermodynamique des systèmes ouverts, la théorie de l'information et la cybernétique. Or, la notion d'information en biologie est loin d'être claire (Godfrey-Smith,

33 Depuis quelques années, le dogme de la biologie moléculaire selon lequel l'ADN est un programme génétique gouvernant l'organisme a commencé à être remis en question. Selon des biologistes comme Jean-Jacques Kupiec (2009), l'expression des gènes serait en grande partie aléatoire. Pour une analyse critique des origines et des implications de la notion de programme génétique, voir Evelyn Fox Keller (2000).

2007; Griffiths, 2001; Sarkar, 2000), notamment à cause de l'attribution de propriétés sémantiques étrangères aux champs de la biochimie et de la biologie, ou encore parce que l'on peut douter qu'elle soit une propriété objective de la matière (Küppers, 1995; Stegmann, 2005). Du point de vue de l'intelligibilité scientifique des phénomènes biologiques, il s'agit d'une stratégie sans doute extrêmement efficace qui réussit à exorciser les fantômes du vitalisme et du finalisme, mais du point de vue de la philosophie, cela ressemble à un pas de côté assez peu satisfaisant.

De plus, la métaphore du programme génétique nous renvoie à celle de l'animal-machine, à ceci près que le modèle de l'ordinateur remplace celui de l'horloge et des automates de Vaucanson. Or, les métaphores du programme et de la machine sont à la fois fructueuses (voir Keller, 2003) et problématiques, y compris pour les partisans du mécanisme, car elles impliquent un créateur, un *design*, un dessein, une finalité. Problème dont la seule issue raisonnable est le recours à l'auto-organisation (qui à cette époque apparaissait sous les traits de la cybernétique) et à la sélection naturelle, c'est-à-dire à un processus « aveugle » qui affaiblit la métaphore initiale du programme.

L'explication des processus téléonomiques du vivant en termes de programme (génétique) pourrait faire une place, au sein de l'approche systémique, à la normativité du langage téléologique et fonctionnel. En effet, des erreurs s'introduisent parfois lors de la copie, de la transcription ou de l'exécution d'un programme, ou existent déjà en son sein. Ces erreurs sont des dysfonctions, mais elles peuvent aussi être à l'origine de nouvelles fonctions ou de l'amélioration de celles existantes. Qu'est-ce qui détermine le caractère positif (bénéfique) ou négatif (délétère) de ces « erreurs » ? Dans un artefact, c'est l'intérêt des utilisateurs et pas le programme lui-même qui en détermine la valeur. Dans quelle mesure peut-on parler d'intérêts à propos des organismes biologiques et à quoi les attribuerait-on : aux gènes, aux individus, aux espèces… ? Au lieu de les considérer comme des erreurs — par rapport à une norme préétablie (laquelle ?) — on pourrait y voir de simples variations, laissant ainsi à la sélection naturelle le loisir d'établir cette norme. Est-ce que la sélection naturelle peut établir une norme de bonne programmation ou de bon fonctionnement d'un système, ou se limite-t-elle à dire lesquels fonctionnent mieux — en termes de succès reproductif — dans un environnement donné ?

Il est loin d'être évident qu'une stratégie centrée sur la notion de programme soit en mesure de justifier la normativité du langage téléologique en biologie. La question est : qui détermine la norme — et comment ? Est-ce le programme lui-même ? Est-ce la sélection naturelle ? Est-ce l'organisme, la population, l'espèce… ? Ou est-ce l'investigateur qui, eu égard à ce qu'il interprète comme étant les intérêts de l'organisme ou de l'espèce (survie, reproduction), eu égard au travail de la sélection naturelle, et à la lumière des régularités statistiques (Boorse, 1977),

assigne une valeur normative aux « erreurs » de ce qu'il décrit — métaphoriquement ou pas (Godfrey-Smith, 2007) – comme étant un « programme » génétique ?

2. Conception cybernétique

La théorie cybernétique a été formulée tout d'abord par Arturo Rosenblueth, Norbert Wiener et Julian Bigelow (1943) puis développée de manière plus approfondie par Wiener (1948), Braithwaite (1953), Ashby (1956) et d'autres. La proposition de Nagel, que nous verrons plus loin, reprend et modifie celle de Gerd Sommerhoff (1950) qui cherchait à déterminer les conditions générales qu'un système doit satisfaire pour être considéré comme « *goal-directed* » indépendamment de sa nature (humaine, biologique, artificielle).

La cybernétique n'établit aucune distinction de principe entre le comportement humain et celui d'un autre système quelconque, qu'il soit naturel ou artificiel. Dans leur article fondateur, Rosenblueth, Wiener et Bigelow (1943) affirmaient en effet ne s'intéresser qu'au comportement *externe* des systèmes, laissant de côté la question de leur organisation interne et le problème de l'intentionnalité. Les états mentaux sont d'ailleurs considérés comme non observables et ne relevant pas de la démarche scientifique (Rosenblueth & Wiener, 1950). La cybernétique est définie par eux comme l'étude des mécanismes téléologiques ou des systèmes dont la finalité (*purpose*) se manifeste par l'ajustement dynamique de leur comportement aux variations environnementales en vue de poursuivre un objectif donné (*goal*). La thèse principale des auteurs est que ce comportement téléologique repose essentiellement sur des boucles de rétro-alimentation négative (*negative feed-back*) et peut être défini en ces termes[34].

Mais cette approche externaliste et comportementaliste est également compatible avec l'idée d'une représentation interne de l'objectif à atteindre, car les systèmes téléologiques sont contrôlés par ce que Thomas Simon (1976, p. 60) appelle un évaluateur interne. L'idée est que, dans un système téléologique, la détermination du comportement est quelque chose d'interne au système. Un évaluateur interne, dit Simon, est « une opération à l'intérieur du système qui mesure d'une manière ou d'une autre les sorties (*outputs*) conformément à un ensemble de règles et qui, par ce biais, provoque des changements dans d'autres variables du système ». Dans le diagramme ci-dessous (Fig. 5), l'évaluateur interne est un comparateur qui maintient les valeurs de sortie du système à l'intérieur d'une fourchette spécifiée par le système lui-même. C'est le type le

34 Cette approche comportementaliste a rapidement été critiquée pour son réductionnisme (Taylor, 1950a, 1950b).

plus simple de système contrôlé par *feedback* (thermostat, régulateur à boules de Watt, etc.).

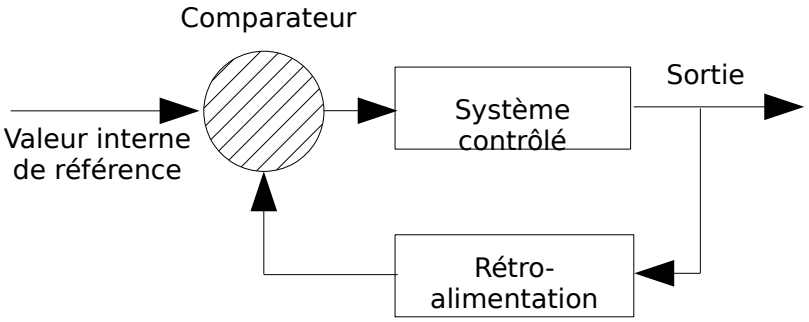

Figure 5: Système cybernétique simple.

Un missile autoguidé, par exemple, contrôle sa trajectoire en comparant régulièrement sa position et sa direction avec celles de la cible pour en compenser les écarts. Si la cible est fixe, les valeurs de référence peuvent être préalablement stockées dans le système. Si elle est mobile, l'autodirecteur du missile doit les obtenir par détection infrarouge ou radar. Il existe des systèmes plus complexes où l'évaluateur interne n'est pas nécessairement un comparateur et où des sous-systèmes d'ordre supérieur modifient les évaluateurs internes des sous-systèmes d'ordre inférieur. De tels systèmes peuvent notamment changer d'objectif après un certain nombre d'essais infructueux. D'après Simon (1976, p. 63), ces systèmes hiérarchiques sont capables d'effectuer des choix et font preuve d'un comportement intentionnel (*purposeful goal-directed activity*).

La gamme des systèmes téléologiques (ou téléonomiques) est donc plus ou moins large en fonction de l'approche que l'on choisit d'adopter. Pour certains, elle se limite aux êtres humains et aux agents intentionnels, c'est-à-dire aux entités dotées d'attitudes propositionnelles qui sont capables de se représenter la finalité de leurs actions. Pour d'autres, elle inclut aussi les systèmes issus de la sélection naturelle, ou bien ceux qui obéissent à un programme, ou, plus généralement, ceux qui manifestent un certain type de comportement. Dans tous les cas, il semble légitime d'attribuer une finalité aux systèmes concernés dans la mesure où les fins ne sont pas incompatibles avec les principes et les connaissances des sciences modernes, et dans la mesure où cette attribution permet d'expliquer le comportement du système dans une description de type « X fait Y pour Z ».

Cependant, il nous semble que Thomas Simon propose là une traduction des explications téléologiques qui substitue une notion problématique par une autre : la notion de but fait place à celle d'évaluateur interne qui remplace les boucles de rétro-alimentation négative de Wiener. Au lieu de dire que « le système *S* réalise l'activité *A* à cause du

but *B* », la formulation cybernétique révisée par ses soins affirme que « le système *S* réalise l'activité *A* à cause des opérations que réalise l'évaluateur interne de *S* » (Simon, 1976, p. 64). À travers cette formulation elliptique, Simon veut montrer que les buts peuvent avoir une efficacité causale dans la mesure où ils sont utilisés pour évaluer et contrôler l'activité du système et que, par conséquent, les explications téléologiques sont scientifiquement légitimes.

Ce qui différencie sa proposition de celle des autres auteurs est la manière dont il considère les notions téléologiques. Ces dernières, dit-il, peuvent être conçues comme des termes théoriques. En tant que telles, elles n'ont pas besoin d'être éliminables en faveur d'une description mécanique. Par exemple, si l'on ne sait pas comment fonctionne le mécanisme interne d'un missile autoguidé, il est selon lui légitime de postuler l'existence d'une structure téléologique interne comme entité théorique pour expliquer son comportement.

Ernest Nagel, de par son influence, est un auteur clef dans le débat sur la téléologie et les fonctions. Dans *The Structure of Science* (1961), il cherche à garantir l'unité des sciences en montrant que les explications téléologiques en biologie ne sont pas incompatibles ni véritablement différentes des explications mécanistes. Dans cette perspective, il cherche à montrer qu'elles n'assurent pas l'autonomie de la biologie vis-à-vis de la physicochimie, car on peut les traduire en termes non téléologiques. Cela le conduit à proposer une traduction réductive des attributions fonctionnelles qui tient compte à la fois de l'organisation du système et de son environnement et omet toute référence à l'histoire ou à l'origine causale de l'item auquel on attribue la fonction. Sa définition est la suivante :

> [A] teleological statement of the form "The function of *A* in a system *S* with organization *C* is to enable *S* in environment *E* to engage in process *P*" can be reformulated more explicitly by: Every system with organization *C* and in environment *E* engages in process *P*; if *S* with organization *C* and in environment *E* does not have *A*, then *S* does not engage in *P*; hence, *S* with organization *C* must have *A*. (Nagel, 1961, p. 403).

En inscrivant cette réduction dans le cadre du modèle explicatif Déductif-Nomologique (D-N), il transforme les fonctions en conditions nécessaires de l'effet recherché. Par exemple, dire que la fonction de la chlorophylle dans les plantes est de leur permettre de réaliser la photosynthèse revient à affirmer que la présence de chlorophylle dans les plantes est une condition nécessaire à la photosynthèse. Or, cette définition ne résiste pas à un contre-exemple : la fonction du cœur est de pomper le

sang bien que la présence du cœur ne soit pas nécessaire pour réaliser cette tâche puisqu'il existe des cœurs artificiels (Cummins, 1975, p. 51)[35].

Le modèle D-N qu'applique Nagel ne donne pas une explication causale de la présence de l'item[36], il explique de manière non causale l'un des rôles que cet item joue dans un système donné. En privant les attributions fonctionnelles de toute référence causale, Nagel se débarrasse efficacement du problème des causes finales, mais aussi de ce qui distingue les phénomènes biologiques des autres phénomènes physiques, à savoir, l'apparence de finalité (*goal-directeness*). Car les explications fonctionnelles telles qu'il les interprète au premier abord peuvent en principe être employées aussi bien en anthropologie qu'en chimie ou en géologie (McLaughlin, 2001, p. 73).

Nagel distingue alors deux questions différentes : la première porte sur la structure des systèmes dirigés vers un but (*goal-directed*) et la seconde sur les explications fonctionnelles qui s'y rapportent. S'appuyant sur la cybernétique, Nagel affirme que l'orientation vers un but (*goal-directeness*) n'est pas un comportement spécifique aux êtres vivants. Deux des principales caractéristiques du comportement de ces systèmes sont la plasticité des processus (le but peut être atteint de manières différentes, à partir de situations initiales différentes) et la persistance des processus (résistance aux perturbations). Exemple : le sang contient une concentration stable de 90% d'eau malgré les variations auxquelles le corps est soumis. La conservation de cette concentration est le but du processus de régulation, et il y a au moins deux variables indépendantes (ou orthogonales) en présence[37]: l'eau retirée du flux sanguin, et l'eau injectée dans le flux sanguin. Il est évident, selon Nagel, que le fait d'être dirigé vers un but (*goal directed*) est une propriété d'un système en vertu de l'organisa-

35 Pour défendre la définition de Nagel, Cummins précise que le cœur est nécessaire dans des circonstances *normales*. Mais ce type de correction, ajoute-t-il, n'est pas suffisant. Une meilleure réponse consisterait à dire que la nécessité du cœur ne fait pas référence à un *token* particulier mais à un type d'organe, soit-il naturel ou artificiel. Nonobstant, la définition de Nagel peut est invalidée autrement par des exemples de multi-réalisabilité et de redondance fonctionnelle.

36 À moins de considérer que les explications du modèle D-N sont causales, à l'instar de Robert Nadeau (1999), ce qui est à la fois légitime et discutable.

37 Pour distinguer un système véritablement « *goal-directed* » d'un système qui tend vers un état d'équilibre (pendule, balle dans un bol, etc.) il faut considérer l'indépendance respective des variables de contrôle. Cette condition, selon Nagel, permet de distinguer les uns des autres sur une base objective: « *It seems, therefore, that the question whether a process is goal-directed can be decided on the "objective" grounds stated in the requirement, rather than on the basis of "subjective" intuitions that often vary from person to person.* » (1977b, p. 274-275).

tion de ses parties (1977b, p. 273) et que cette propriété est indépendante de la nature vivante ou inerte du système (1977b, p. 274).

À partir de l'analyse des systèmes dirigés vers un but, Nagel propose une interprétation des attributions fonctionnelles où la fonction est conçue comme un moyen en vue d'une fin :

> A functional statement of the form: a function of the item i in system S and environment E is F, presupposes (though it may not imply) that S is goal-directed to some goal G, to the realization or maintenance of which F contributes. I will call this account the "goal-supporting" view of biological functions. (Nagel, 1977a, p. 297)

On peut aussi voir dans la conception de Nagel une relation un peu plus complexe où l'item fonctionnel est un moyen en vue d'une fin (sa fonction) qui à son tour est le moyen d'une fin supérieure, à savoir, le but vers lequel est orienté le système (McLaughlin, 2001, p. 75).

Sur ce point, Nagel et Hempel tombent d'accord. Chez ce dernier, le contenu d'une affirmation comme « La fonction des battements du cœur chez les vertébrés est la circulation du sang » est rendu plus explicite de la manière suivante: « *Les battements cardiaques ont pour effet de faire circuler le sang, ce qui assure la satisfaction de certaines conditions qui sont nécessaires pour le fonctionnement correct le l'organisme* ». La relation fonctionnelle est donc là aussi une relation instrumentale où la fonction est relative à une fin. En revanche, ils ne sont pas d'accord sur la caractérisation de la fin en question. Une des critiques de Nagel est que l'idée de « fonctionnement correct » de l'organisme est trop vague pour constituer un critère suffisant afin de distinguer les fonctions des simples effets. Il lui préfère l'idée d'« activité caractéristique » du système tout en confessant la même difficulté que Hempel[38].

Le problème qui se pose est l'identification de la fin (*end*) ou du but (*goal*) relativement auquel se détermine la fonction d'un item. S'il est arbitrairement choisi par l'observateur, alors il est probable que n'importe quel effet d'un item puisse être considéré comme sa fonction. S'il ne dépend pas de l'observateur, peut-on dire qu'il y a un but ou une fin qui soit propre au système ? Selon Peter McLaughlin (2001, p. 76), c'est effectivement le cas. Chez Hempel, le but intrinsèque ultime d'un système serait son propre bien (*welfare*), c'est-à-dire son auto-conservation (*self-maintenance*), tandis que chez Nagel, ce serait son activité caractéristique. Cela soulève trois questions : *Comment reconnaître le bien*

38 Hugh Lehman (1965a) se livre à une critique détaillée de la conception de Nagel et reprend à son compte l'idée de « fonctionnement correct » de l'organisme pour définir les fonctions.

propre ou l'activité caractéristique d'un système ? À quels types de systèmes peut-on attribuer un bien ou un but propres ? Et quelles sont les implications métaphysiques de cette attribution ?

Christopher Boorse défend quant à lui une conception cybernétique similaire à celle de Nagel tout en évitant l'un des principaux écueils de la définition de ce dernier, à savoir, le modèle D-N. D'après lui, « *les fonctions sont, purement et simplement, des contributions à des buts* » (1976, p. 77). Ensuite, tout en distinguant *la* fonction et *une* fonction, il propose pour cette dernière une définition plus précise :

> "A function of X is Z" means that in some contextually definite goal-directed system S, during some contextually definite time interval t, the Z-ing of X falls within some contextually circumscribed class of functions being performed by X during t – that is, causal contributions to a goal G of S. (Boorse, 1976, p. 82)

Selon Boorse, le but d'un système est déterminé par un mécanisme interne qui guide son comportement vers la poursuite de ce but, à la manière des missiles autoguidés. Les êtres vivants, depuis cette perspective, sont des systèmes objectivement dirigés vers plusieurs buts comme par exemple la survie et la reproduction. Faisant remarquer que les chaises et les crayons, auxquels nous attribuons des fonctions, ne sont pas en eux-mêmes des systèmes dirigés vers un but, il propose de considérer leur fonctionnalité du point de vue de la contribution à l'activité des systèmes orientés vers un but que sont les êtres humains. C'est-à-dire que les objets ne disposant pas de mécanismes internes leur permettant d'être eux-mêmes considérés comme des systèmes téléologiques héritent leurs caractéristiques fonctionnelles de l'usage que nous en faisons. Par ailleurs, la diversité des systèmes, des niveaux d'organisation (du gène à l'écosystème) et des intervalles temporels étudiés par les différentes branches de la biologie est responsable d'une pluralité de buts et de fonctions, de sorte que tous les biologistes n'emploient pas le concept de la même façon. Cela conduit l'auteur à proposer ce qu'il appelle une stratégie d'adjectivation consistant à distinguer par le biais de l'adjectif correspondant ses différents usages implicites (2002, p. 72-73) : fonction actuelle, fonction passée, fonction évolutive, fonction évolutive récente, etc.

L'une des principales difficultés de l'approche cybernétique est la discrimination entre les systèmes dirigés vers un but et ceux qui ne le sont pas, car il semble toujours possible de présenter des contre-exemples. On peut ainsi se demander si la cybernétique rend compte d'une distinction pré-existante ou si elle la crée.

Nous avons vu que chez Nagel certains systèmes ont un comportement orienté vers un but, que ce but est une propriété du système liée à l'organisation de ses parties, et que cette organisation doit satisfaire au moins deux conditions : la plasticité et la persistance des processus de contrôle. Autrement dit, les systèmes orientés vers un but sont dotés de mécanismes d'auto-régulation. Chez Hempel, de même, l'analyse fonctionnelle porte sur des systèmes auto-régulés.

Cependant, l'auto-régulation n'est pas nécessaire ni suffisante pour distinguer les systèmes téléologiques ou téliques (ceux auxquels on attribue un but ou une fin) de ceux qui ne le sont pas (Woodfield, 1976). Elle n'est pas nécessaire, car on attribue des buts et des fonctions à des systèmes (manufacturés) dépourvus de mécanismes de régulation, et elle n'est pas nécessaire, car des systèmes auto-régulés ne méritent pas cette attribution. Mark Bedau (1992a) imagine un contre-exemple, sous la forme d'un super-pendule, pour montrer l'incapacité de l'approche de Nagel à distinguer entre les systèmes orientés vers un but et ceux qui tendent vers un état d'équilibre (une bille au fond d'un bol, un pendule, etc.) et auxquels on n'attribue ni buts ni fonctions. Arthur Collins (1978), pour sa part, reconnaît les déficiences de l'approche « béhavioriste » de Nagel et en propose une version corrigée, laquelle est aussitôt invalidée par Peter Achinstein (1978).

Quelles caractéristiques physiques doit avoir un système pour qu'on puisse lui attribuer des fonctions — ou plutôt en attribuer à ses parties ? Nagel tente de répondre à cette question en identifiant ces systèmes comme « *goal-directed* » et en proposant une théorie générale inspirée de la cybernétique qui lui évite de distinguer les êtres vivants des autres systèmes physiques. Tous les organismes biologiques sont vraisemblablement des systèmes auto-régulés, quelques artefacts complexes le sont aussi, de même que certains systèmes naturels inanimés. Dans les étoiles plusieurs mécanismes de rétroaction contribuent à conserver leur diamètre stable autour d'une valeur d'équilibre sur un principe analogue au régulateur à boules de Watt. On pourrait dire par exemple que l'une des fonctions des réactions de fusion nucléaire est de produire l'énergie nécessaire pour compenser la compression gravitationnelle, maintenir constant le diamètre de l'étoile et assurer la continuation de son « activité caractéristique ». C'est une explication téléologique inoffensive d'un point de vue métaphysique et aisément traduisible en termes non téléologiques. Tellement inoffensive qu'on ne peut pas la prendre au sérieux, contrairement aux explications téléologiques que l'on rencontre en biologie ou dans le domaine des artefacts. Aucun astrophysicien ne serait prêt à parler téléologiquement des réactions nucléaires d'une étoile avec la même conviction qu'un biologiste à propos de la photosynthèse ou des mitochondries. Pourquoi ? Quelles caractéristiques physiques doit avoir un système pour qu'on puisse *sérieusement* lui attribuer des fonctions ?

Nagel ne répond pas à cette question[39]. Mais conformément à l'intuition de Hempel — qui ne prend pas les explications téléologiques au sérieux — il semble que nous n'attribuions sérieusement de fonctions à un système ou à ses parties que lorsque nous reconnaissons à l'activité fonctionnelle un intérêt ou un bénéfice pour le système lui-même ou pour un autre. Or, il semble également que cette reconnaissance soit attachée principalement — si ce n'est exclusivement — aux organismes biologiques et aux artefacts. Les premiers ayants des intérêts propres (survie et reproduction) et les seconds un intérêt pour autrui (pour leur créateur ou leur utilisateur). Par conséquent, la question précédente pourrait/devrait être : Quelles caractéristiques physiques doit avoir un système pour qu'on puisse lui attribuer un *intérêt* ou un *bénéfice* ?

Cette distinction entre systèmes sur une base valorative rejoint celle qu'on peut établir sur une base normative. En effet, aucun astrophysicien sérieux n'affirmerait que les processus à l'œuvre dans une étoile dysfonctionnent, alors qu'on le fait habituellement pour les êtres vivants et les artefacts. Pourquoi ? Quelles caractéristiques physiques doit avoir un système pour qu'on puisse lui attribuer une *dysfonction*[40] ?

Le problème de ces questions est quelles présupposent toutes une partie de la réponse, à savoir, que la différence se trouve dans les caractéristiques physiques des systèmes. Nous disions plus haut que selon Nagel (1977b, p. 273) le fait d'être dirigé vers un but (*goal-directedness*) est une propriété du système en vertu de l'organisation de ses parties. La question qu'il se pose est presque d'ordre scientifique, et c'est ainsi qu'il y répond, conformément à sa posture naturaliste. Face aux explications vitalistes et finalistes de la téléologie biologique, Nagel considère le comportement et l'organisation des êtres vivants comme des manifestations d'un phénomène naturel plus général dont il faut arriver à comprendre les causes et les mécanismes dans le cadre des sciences de la nature. Et il ne fait aucun doute que le comportement des êtres vivants, aussi complexe soit-il, est en principe explicable en termes naturalistes. Mais Nagel ne rend pas compte des dimensions valorative et normative des explications téléologiques ni du fait qu'elles sont attachées à certains

39 Lorsqu'il évoque un surplus de signification (*surplus meaning*), qu'il interprète en termes d'emphase et de perspective dans la formulation, Nagel (1961, p. 421-422) compare les discours téléologique et non-téléologique, mais il ne rend pas compte des différences qui existent entre les attributions fonctionnelles portant sur des systèmes « *goal directed* ».

40 En formulant une objection contre l'assimilation par Ruse des fonctions à des adaptations, Nagel (1977a, p. 298) cite l'exemple des gènes qui produisent la couleur jaune de certains oignons : bien que cette couleur n'ait aucune valeur adaptative, dit-il, les gènes en question ont bien la fonction correspondante. Pourtant, si pour une raison ou pour une autre l'oignon n'était pas en mesure de produire cette couleur jaune, sans que cela ait la moindre incidence négative sur sa *fitness,* parlerait-on de dysfonction ?

systèmes (biologiques et techniques) et pas à d'autres. Et on peut douter que ces dimensions valorative et normative soient éclaircies par une investigation empirique sur les caractéristiques physiques des systèmes en question. C'est-à-dire que l'on peut douter de ce que ces dimensions soient seulement relatives aux systèmes physiques et pas aussi à l'observateur.

Boorse (2002, p. 75-77) répond partiellement aux questions que nous nous posons à propos de l'approche de Nagel à travers sa réponse à l'objection suivante : Si les mécanismes d'autorégulation qui maintiennent une concentration stable de 90% d'eau dans le sang subissaient une modification telle que leur but passait à être de seulement 70%, alors il n'y aurait pas de raison, du point de vue de l'approche cybernétique, pour considérer que ce nouveau but est « incorrect » ou « dysfonctionnel » bien que ses conséquences pour l'organisme soient manifestement délétères. La seule réponse possible, continue la critique, serait de dire que le nouveau but est dysfonctionnel dans la mesure où il ne contribue pas au but d'un système de niveau supérieur, à savoir, en dernière instance, à la survie et à la reproduction de l'organisme. Or, faire référence à la vie et à la santé du système de plus haut niveau pour justifier l'attribution de buts à ses sous-systèmes romprait l'unité et la symétrie de l'approche cybernétique : il y aurait d'un côté une analyse de la téléologie relative à la survie et la reproduction des systèmes biologiques, et, de l'autre, une analyse de la téléologie des artefacts relative aux intentions humaines.

La réponse de Boorse à cette critique est d'autant plus intéressante pour nous qu'elle rejoint l'une des lignes directrices de notre réflexion, à savoir, l'analyse des fonctions biologiques en parallèle avec l'analyse du vivant. En effet, il s'appuie sur une définition cybernétique de la vie due à Sommerhoff pour dire, succinctement, que les êtres vivants sont juste un type de système naturel *goal-directed* dont le but ultime (*apical goal*) est la continuation de la vie, c'est-à-dire la survie et la reproduction (2002, p. 76).

Par ailleurs, il interprète la dysfonction, et en particulier la maladie, depuis la perspective d'une normalité statistique qui lui évite de prononcer des jugements de valeur (car il défend la neutralité axiologique des attributions fonctionnelles). Une fonction (normale) est selon lui celle exercée normalement (au sens statistique) par un type de système. Cette interprétation l'amène à dire que la fonction d'une boucle de ceinturon pourrait être de dévier les balles si cette circonstance devenait suffisamment fréquente.

Dans le domaine biologique, cette interprétation statistique de la normalité fonctionnelle devrait aussi l'amener à dire qu'au pays des aveugles les borgnes sont dysfonctionnels (puisque statistiquement anormaux). Pour se défendre de cette critique, il développe un curieux argument qui l'amène à dire que la normalité statistique doit inclure une

dimension temporelle pouvant comprendre plusieurs générations, voire des milliers ou des millions d'années :

> « Obviously, some of the species' history must be included in what is species-typical. If the whole earth went black for two days and most human beings could not see anything, it would be absurd to say that vision ceased to be a normal function of the human eye. Actually, any time-slice shorter than a lifetime or two seems too short for the very idea of a species-typical functional design, since identifying many functions in maturation and reproduction requires a longitudinal view of an individual organism and its progeny. Originally, I spoke of including 'milenia' of the species' history, and more recently I said that contemporary Western civilization was 'barely an eye-blink in the history of man'. » (Boorse, 2002, p. 99)

L'argument est curieux pour deux raisons. En premier lieu parce qu'il devrait faire passer sa définition du côté des conceptions étiologiques, à côté de celles de Wright, Neander et Millikan, sans pour autant être une conception darwinienne. En second lieu, l'argument est curieux parce que la fourchette temporelle pertinente pour une attribution fonctionnelle est apparemment arbitraire. L'auteur reconnaît que la fonction normale d'un trait est relative à la période temporelle que l'on considère significative, tout en n'offrant aucun critère sur ce point, et il admet que l'on puisse ne pas savoir, étant donné cette relativité, quelle est la fonction normale du trait (2002, p. 100). Il poursuit en disant que les conceptions rivales (celle de Neander, en l'occurrence) ne font pas mieux. Plus loin (2002, p. 102), il exprime son trouble face aux maladies que la médecine reconnaît comme typiques, voire universelles, que ce soit à l'échelle de toute notre espèce ou seulement d'un groupe d'âge, et affirme qu'il ne peut s'agir que d'une erreur dans la mesure où la médecine ne dispose pas d'un concept cohérent de pathologie qui lui permette de dire qu'une maladie est effectivement universelle. En faisant dépendre son concept de fonction d'une interprétation statistique pour le moins contestable et relative à une période de temps apparemment arbitraire, Boorse limite considérablement la portée et la vraisemblance de sa proposition. C'est peut-être la raison pour laquelle cette proposition n'a guère eu d'écho dans les débats postérieurs sur le concept de fonction.

En ce qui concerne les questions que nous posions à propos de Nagel, on peut inférer que la réponse de Boorse serait la suivante. Pour attribuer une dysfonction à un système, il suffit que l'on puisse lui attribuer une fonction, car les dysfonctions et les pathologies n'ont en effet pour lui qu'une signification statistique par rapport à l'exercice normal ou typique d'un trait (ce que Nagel appelle son activité caractéristique). Cette définition statistique lui permet donc de défendre la neutralité axiologique des attributions fonctionnelles et, par conséquent, de ne pas répondre à notre question concernant les caractéristiques physiques que

doit avoir un système pour qu'on puisse lui attribuer un intérêt ou un bénéfice. Le problème est que, dans ces conditions, Boorse n'est pas en mesure de répondre à la première question portant sur les caractéristiques que doit avoir un système pour qu'on puisse *sérieusement* attribuer des fonctions à ses parties. Si la notion de but n'a aucune connotation valorative, et si le but d'un système est déterminé seulement par ses mécanismes de compensation, alors la définition de Boorse est applicable à n'importe quel système naturel présentant de tels mécanismes. Elle est donc vulnérable, comme celle de Nagel, au contre-exemple de l'étoile que nous avons présenté plus haut. Or, si l'on peut attribuer un but à n'importe quel système naturel abiotique présentant certains mécanismes de compensation, alors il n'y a aucune différence entre l'attribution d'un but et l'attribution d'un état d'équilibre dynamique. Pourtant, aucune science naturelle en dehors de la biologie ne considère l'état d'équilibre vers lequel tend un système, si complexe soit-il, comme étant un but de ce système ; et aucune n'attribue de fonctions à ses parties. Par conséquent, la définition de Boorse, comme celle de Nagel, revient à dire que le langage fonctionnel et téléologique n'est qu'une façon de parler dont les biologistes pourraient facilement se passer puisqu'il n'apporte rien par rapport au discours des autres sciences naturelles.

3. Conception dispositionnelle

Comme Nagel et contrairement à Wright, Robert Cummins (1975) considère que la fonction d'un organe n'a rien à voir avec les raisons de son existence, et que les effets d'un élément fonctionnel dans un système ne sont pas causalement pertinents pour expliquer sa présence. Il essaie de montrer deux choses :

(1) que la fonction n'explique pas la présence d'un trait et

(2) qu'elle n'est pas un effet de la présence de ce trait.

Contre (1) il affirme que la fonction ne détermine pas la chose qui la remplit. Puisque deux items ou deux traits peuvent remplir la même fonction, celle-ci n'explique pas pourquoi un seul des deux existe. De plus, mentionner la fonction de quelque chose ne répond à la question de sa présence que dans le cas des artefacts. En effet, tandis que pour ceux-ci il y a toujours une raison expliquant l'existence de leurs différents éléments (l'intention du créateur), nous ne pouvons invoquer, concernant les objets biologiques, aucune raison particulière, un élément fonctionnel pouvant être là du fait d'une mutation aléatoire du code génétique. La fonction n'explique donc pas la présence d'un trait, mais les capacités de l'organisme en vertu desquelles il survit.

Contre (2) il montre que nous sommes incapables d'identifier et de distinguer correctement les effets fonctionnels de ceux qui ne le sont pas, d'autant plus que leur caractère fonctionnel est relatif à un environnement. Il critique notamment la tentative de définir les effets fonctionnels comme étant ceux qui contribuent à la survie de l'espèce, car le vol d'un oiseau, par exemple, est une capacité qui demande une analyse en termes de fonctions indépendamment de sa valeur évolutive.

Cummins propose une conception alternative déconnectant l'analyse fonctionnelle du langage téléologique dont il fait par ailleurs une critique virulente (2002a). Il définit la fonction d'un item à partir du rôle qu'il joue dans le système auquel il appartient. Plus précisément, la fonction d'un item est sa capacité à contribuer à une capacité du système auquel il appartient. Cette définition rejoint celle formulée six ans plus tôt par Morton Beckner (1969), à ceci près ce dernier ne parlait pas de capacités mais d'activités.

D'après Cummins, une fonction n'est pas un effet, mais une disposition — les dispositions étant des propriétés qui impliquent une régularité de type nomologique : quand une certaine condition est remplie, la disposition se réalise — et les attributions de fonctions s'inscrivent dans le cadre d'une analyse fonctionnelle des capacités d'un système. Comme chez Nagel, les fonctions s'inscrivent dans une relation de type moyens–fin indépendante du système étudié. Mais tandis que Nagel cherche à subsumer un certain type de phénomènes sous des lois générales qui l'expliquent (dans le cadre du modèle de couverture légale), Cummins adopte une stratégie consistant à décomposer une disposition d'un système particulier en dispositions plus simples de ses éléments constitutifs, sans recourir à des lois générales.

L'analyse fonctionnelle consiste à expliquer les dispositions ou capacités d'un système en analysant le rôle de chacun des sous-systèmes qui le composent, comme on le fait pour une chaîne de montage ou un circuit électronique. Un organisme biologique, de même qu'une machine, est ainsi considéré comme un ensemble de systèmes dotés de certaines capacités et analysables à leur tour en termes de capacités des sous-systèmes. Les fonctions qu'on attribue aux traits d'un organisme correspondent par conséquent à la capacité de ces traits à contribuer à une capacité plus générale de l'organisme :

> "*x* functions as a ϕ in *s* (or: the function of *x* in *s* is to ϕ) relative to an analytical account *A* of *s*'s capacity to ψ just in case *x* is capable of ϕ-ing in *s* and *A* appropiately and adequately accounts for *s*'s capacity to ψ by, in part, appealing to the capacity of *x* to ϕ in *s*." (Cummins, 1975, p. 762)

Ou plus simplement :

> "When a capacity of a containing system is appropriately explained by analyzing it into a number of other capacities whose programmed exercise yields a manifestation of the analyzed capacity, the analyzing capacities emerge as functions." (Cummins, 1975, p. 765)

La fonction d'un trait ou d'un item est son rôle causal ou sa contribution à une capacité du système. C'est une théorie générale indépendante de la nature du système qu'on étudie ; elle s'applique aussi bien à la physiologie qu'à la psychologie, aux sciences sociales ou à l'ingénierie. Elle ne rend pas compte de la signification du concept de fonction mais d'un style analytique d'explication qui s'exprime fonctionnellement.[41] L'explication en question ne porte pas sur l'item auquel on attribue la fonction mais sur le système auquel il appartient.

N'importe quelle capacité d'un système est analysable fonctionnellement, avec pour seule restriction la pertinence de l'explication. Il serait par exemple futile — quoique possible — de vouloir expliquer les capacités sonores du système circulatoire des mammifères en affirmant que les battements du cœur ont pour fonction de produire un bruit régulier. L'intérêt de l'analyse fonctionnelle est proportionnel à la complexité relative et à la différence entre l'*explanans* et l'*explanandum* ; or le bruit du cœur n'est ni différent ni moins « sophistiqué » que le bruit du système circulatoire, il en fait simplement partie (Cummins, 1975, p. 764). Le problème de l'analyse fonctionnelle en général et de cette restriction en particulier est qu'elles rendent paradoxalement difficile l'attribution de fonctions à des objets aussi simples qu'une cuillère.

Trois critiques principales ont été formulées contre la proposition de Cummins. La première concerne son manque d'engagement ontologique. Les deux autres consistent à dire qu'elle attribue des fonctions là où il n'y en a pas (trop grande libéralité ou « promiscuité » de la théorie), et qu'elle n'en attribue pas là où il y en a (absence de normativité). Pour une défense des fonctions entendues comme rôle-causal, voir Amundson & Lauder (1994).

---

41 « However, it is the analytical style of explanation, especially as applied to complex capacities, that interests me, not the proper explication of the concept of *function*. Thus 'functional analysis' can be understood here as no more than a technical term for a theory designed to explain a capacity or disposition via property analysis. » Cummins, *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press, 1983, p. 195, cité par McLaughlin (2001, p. 121).

Tandis que chez Nagel le but vers lequel le comportement d'un système est cybernétiquement dirigé constitue une propriété de ce système, la capacité du système faisant l'objet d'une analyse fonctionnelle suivant l'approche de Cummins peut être n'importe quelle capacité librement choisie par l'observateur selon ses propres intérêts épistémiques. Pour cette raison, plusieurs auteurs (Bigelow & Pargetter, 1987; Enç & Adams, 1992) interprètent — et rejettent — sa posture comme une forme d'éliminativisme.[42] Nous verrons que Paul Davies (2001), qui défend et développe l'approche de Cummins, rejette également son éliminativisme et adopte une posture réaliste à propos des fonctions.

Au-delà de la capacité du système, c'est aussi le choix du niveau d'analyse — dont dépend directement l'attribution des fonctions — qui est laissé à l'arbitre de l'observateur, ainsi que le choix des conditions extérieures (environnementales par exemple) qui déterminent la possibilité d'une capacité donnée : une plante n'est pas capable de réaliser la photosynthèse dans une chambre noire, ni dans une atmosphère sans dioxyde de carbone ; une cafetière électrique n'est pas capable de faire le café sans une source d'électricité, sans café moulu, et sans quelqu'un pour la mettre en marche.

La seconde critique concerne le manque de normativité de l'analyse fonctionnelle. Puisque seule des dispositions actuelles sont causalement pertinentes pour expliquer les capacités d'un système, il n'est pas question de dysfonctions ni de malfonctions.[43] Un cœur malformé, incapable de pomper le sang, ne dysfonctionne pas ; il ne possède tout simplement pas la fonction correspondante. Cette conséquence de l'analyse fonctionnelle est au moins aussi choquante que l'éliminativisme, car elle consiste à nier toute valeur aux concepts de maladie ou de mort dans la mesure où ils impliquent une malfonction ou une dysfonction.

Curieusement, trois des modèles d'analyse fonctionnelle que cite Cummins sont l'ingénierie (chaînes de montage, circuits électroniques), la physiologie et la psychologie. Or, dans ces trois domaines, le langage normatif n'est pas exclu, bien au contraire.[44]

---

42 « [Eliminativist] views identify functions merely with the activities (or dispositions) of a character that happen to interest the investigator, and they effectively reject the intuition that a real difference exists between dispositions and functions. » (Enç & Adams, 1992, p. 637)

43 Valerie G. Hardcastle (2002) soutient qu'une approche pragmatique comme celle de Cummins peut rendre compte de la normativité des fonctions, mais elle n'explique pas comment. Au contraire, pour Röhl et Jansen (2014), les dispositions sont incompatibles avec une conception normative des fonctions : « *We argue that functions should not be taken as a subtype of dispositions. The strongest reason for this is that any view that identifies functions with certain dispositions cannot account for malfunctioning, which is having a function but lacking the matching disposition.* »

Les machines et les systèmes électroniques, comme les organismes et les esprits, fonctionnent bien ou mal et parfois dysfonctionnent complètement. C'est la raison pour laquelle il existe des techniciens et des médecins qui s'occupent de les réparer. Prenons une chaîne de montage automobile et court-circuitons l'une des machines qui la composent, de sorte que l'ensemble de la chaîne s'arrête et perde sa capacité à fabriquer des voitures. Les ouvriers doivent-ils abandonner l'usine en se disant que les machines n'ont plus de fonctions (du point de vue de la capacité de la chaîne à fabriquer des voitures) et qu'eux-mêmes n'ont plus de travail ? Ne devraient-ils pas plutôt attribuer l'arrêt de la chaîne au dysfonctionnement d'une machine, procéder aux réparations pertinentes et se remettre au travail ? Dans ce second cas, il faudrait qu'ils attribuent à la machine cassée une fonction que d'après Cummins elle ne devrait pas avoir. L'analyse fonctionnelle telle qu'il la présente ne rend pas compte de cette normativité des attributions fonctionnelles qui, pourtant, est inévitablement attachée à l'analyse des systèmes techniques. À moins de penser que la fonction d'un artefact est relative aux intentions des créateurs ou de ses utilisateurs. Dans ce cas, les ouvriers continueraient effectivement à reconnaître la fonction des machines après que le système dont elles font partie ait perdu sa capacité à produire des voitures. Mais cette solution est insatisfaisante. D'abord parce qu'elle crée une distinction injustifiée entre les fonctions des artefacts et celles des autres types de systèmes. Ensuite parce qu'elle rend l'attribution de fonctions indépendante de l'organisation du système. Et enfin parce qu'elle ne rend pas compte de l'attribution de fonctions à des artefacts pour lesquels on ignore les intentions qui ont guidé leur création et leur utilisation.

Lorsqu'ils ont découvert le mécanisme d'Anticythère (Fig. 6), les archéologues ont attribué à ses parties des fonctions — alors indéterminées — et au système dans son ensemble une capacité — tout aussi inconnue — qu'il ne possède pas actuellement. Cette attribution s'appuie sur plusieurs suppositions. La première est que cet amas d'engrenages rouillés correspond aux restes détériorés d'un objet manufacturé. La seconde est que l'organisation des éléments de cet objet n'est pas gratuite, qu'elle est le produit d'un *design*. La troisième est que son organisation dotait l'objet original de capacités particulières qui ont motivé sa construction. Les fonctions qu'on attribue aux fragments de l'objet

44 Ruth G. Millikan souligne fort à propos que les exemples qu'emploie Robert Cummins pour illustrer son analyse fonctionnelle correspondent davantage à des types de systèmes qu'à des systèmes matériels concrets, ce qui lui permettrait de justifier un langage normatif : « *In describing the general form that functional analyses take, Cummins mentions flow charts, circuit diagrams, and computer programs. Notice that representations of this kind generally specify ideal rather than actual systems. The circuit diagram that comes with your washing machine represents how it was designed or intended to function, not necessarily how it does function* » (2002, p. 119).

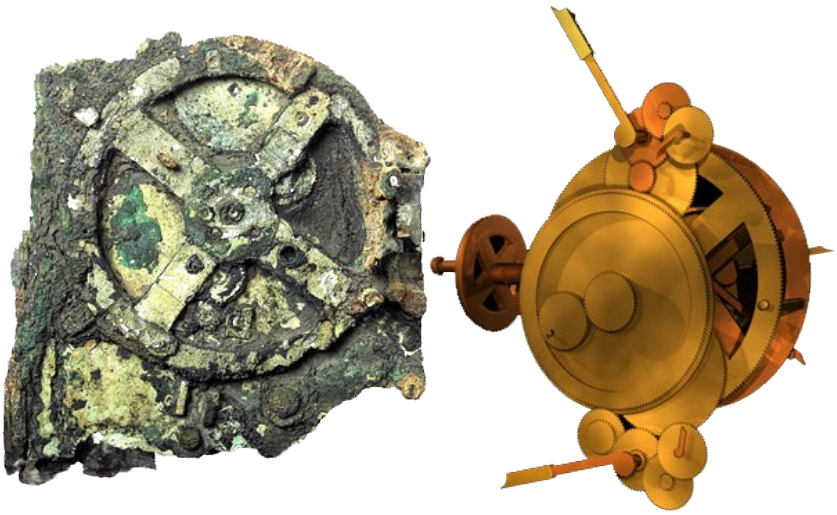

Figure 6: Mécanisme d'Anticythère. À gauche, le fragment principal, vu de face. À droite, une recréation idéale du mécanisme.

découvert au large de l'île d'Anticythère correspondent par conséquent à celles d'un autre objet (l'artefact original) auquel on attribue des capacités qui expliquent à la fois son existence et son organisation. Ces capacités ne sont pas relatives aux intérêts épistémiques des archéologues actuels mais aux intérêts pratiques ou scientifiques des utilisateurs de l'époque.

Pour déterminer quelles étaient ces capacités, il faut faire correspondre deux hypothèses, l'une porte sur les intentions et les compétences des créateurs potentiels de l'artefact, et l'autre sur sa structure originale. Ces hypothèses reviennent à classer l'objet dans une catégorie fonctionnelle (horloge, astrolabe, machine à calculer[45]) et à en déduire un *type* d'organisation qui s'ajuste aux fragments d'engrenages retrouvés. À partir de là, on peut appliquer une analyse fonctionnelle à la Cummins ; mais il convient de noter que cette analyse porte maintenant sur un *type* de système dont la machine d'Anticythère est peut-être le seul *token*[46]. Bien qu'ils n'en aient plus la disposition, roues, aiguilles et cadrans conservent néanmoins la fonction qui correspond à leur type. Et c'est relativement à ce type que l'on peut juger le système actuel *(token)* en termes normatifs.

45 Certains chercheurs (Freeth, 2008; Freeth et al., 2006) affirment que le mécanisme d'Anticythère est un calculateur astronomique extrêmement complexe pouvant prédire le mouvement du Soleil, de la Lune et des planètes autour de la Terre, ainsi que les dates des éclipses et le calendrier des Olympiades. Voir Fig. 6 (Marchant, 2006).

46 Une telle interprétation est suggérée par Neander (2010, p. 102), qui signale par ailleurs le problème de distinguer, en parlant de types, entre les capacités qui sont des fonctions et celles qui ne le sont pas (2012a).

La troisième critique formulée contre l'approche de Cummins consiste à dire qu'elle attribue des fonctions là où il n'y en a pas[47]. Les contre-exemples sont de deux types : d'un côté, des systèmes naturels abiotiques ; de l'autre, des systèmes biologiques dont on analyse une capacité négative, comme celle de mourir d'une infection du foie ou de développer un cancer. En effet, malgré leur rôle causal indéniable dans le déclenchement des pathologies cancéreuses, nous n'attribuons pas aux oncogènes la fonction correspondante[48]. De manière générale, il semble que les fonctions ont toujours une valeur positive, que ce soit pour le système lui-même ou pour quelqu'un qui lui est associé (créateur, utilisateur). Or, l'approche de Cummins ne rend pas compte de cette dimension valorative des attributions fonctionnelles.

Pourquoi ne peut-on pas dire que la fonction de l'eau est de réfracter la lumière et que celle des gouttes en suspension dans l'air est de produire des arcs-en-ciel (Bigelow & Pargetter, 1987, p. 184) ? En réalité on peut, mais ces attributions sont inutiles et gênantes. Inutiles parce qu'on peut s'en passer, et gênantes soit parce qu'elles donnent la fausse impression d'être des explications téléologiques, soit parce que ce sont des explications téléologiques fausses. Tous les systèmes et les phénomènes physiques sont explicables indépendamment du concept de fonction alors que les systèmes techniques, biologiques, psychologiques et sociaux ne le sont pas. Or l'analyse fonctionnelle de Cummins s'applique indifféremment à tous les domaines et ne rend pas compte de cette distinction. La question est la suivante : Le fait d'attribuer une fonction à un item apporte-t-il quelque chose de plus par rapport au fait de lui attribuer un rôle causal ou une disposition ? Et si oui, pourquoi ce « surplus de signification » — comme l'appelle Nagel — des attributions fonctionnelles semble-t-il être limité à certains domaines d'objets ?

Cummins ne s'intéresse pas aux raisons de la présence des éléments des systèmes dont il fait l'analyse. Il se contente d'expliquer comment ils fonctionnent sans se demander pourquoi il fonctionnent ainsi ni pourquoi les systèmes sont constitués comme ils le sont. Ces deux dernières questions trouvent en physique une réponse qui oscille entre le hasard et la nécessité : les systèmes physiques sont le fruit de lois universelles nécessaires et de phénomènes aléatoires qui généralement obéissent à leur tour à d'autres lois — statistiques — nécessaires. En revanche, l'organisation et le fonctionnement des systèmes biologiques, techniques, psychologiques et sociaux, ne sont ni aléatoires ni nécessaires. Et tandis que la

47 Boorse (2002, p. 65) recueille plusieurs contre-exemples de la littérature contre les fonctions de Cummins. Davies (2001, p. 75) y ajoute celui formulé par M. Matthen (1988, p. 15).

48 Les oncogènes ont normalement une fonction d'activateurs du cycle cellulaire dans un tissu donné et à un moment précis de la différenciation ou du développement.

structure d'un système physique s'explique seulement à la lumière de ses causes, celle d'un système biologique ou technique semble s'expliquer aussi à la lumière de ses conséquences ou, pour le dire à la manière de Cummins, à la lumière des capacités qu'elle rend possibles. C'est-à-dire que certaines des capacités d'un tel système — et elles seulement — semblent justifier son organisation. Si les engrenages du mécanisme d'Anticythère étaient disposés autrement, il ne serait pas capable de calculer le mouvement des planètes. Par contre, il serait toujours capable d'accumuler la poussière, de couler au fond de l'eau, et de servir comme presse-papiers. Ce n'est qu'en le considérant comme un calculateur astronomique que l'on répondra de manière satisfaisante à la question « Pourquoi possède-t-il tels et tels engrenages disposés de telle et telle façon ? » De même, ce n'est qu'en considérant le cœur comme une pompe du système circulatoire que l'on comprendra pourquoi il réalise continuellement des contractions rythmiques, pourquoi il est creux, composé de plusieurs cavités, connecté à des vaisseaux sanguins, etc.

Depuis cette perspective, les fonctions qu'on attribue aux éléments d'un système ne rendent pas compte d'une capacité arbitrairement choisie par l'investigateur suivant ses propres intérêts épistémiques. Elles rendent compte d'une capacité particulière du système qui est la conséquence et la raison d'être de son organisation. Ce qui nous frappe chez les êtres vivants et dans certains artefacts, c'est précisément cet ordre, cette organisation que les systèmes naturels abiotiques n'ont pas. C'est lui qui pousse les créationnistes à croire qu'ils sont le produit d'un créateur intelligent. Et c'est lui qui pousse les archéologues à penser que l'amas de fragments métalliques rouillés retrouvé au large d'Anticythère est le vestige dysfonctionnel d'une machine dotée de capacités extraordinaires. Les fonctions des éléments du mécanisme d'Anticythère sont relatives à ces capacités-là, pas à n'importe quelle autre.

4. Conception dispositionnelle–hiérarchique

Paul Sheldon Davies (2001) défend l'idée que les fonctions ne sont rien d'autre que des capacités systémiques qui contribuent à la réalisation de capacités de plus haut niveau que l'on veut comprendre et contrôler. Il propose d'améliorer l'approche systémique de Cummins en y ajoutant quatre thèses :

(1) L'approche systémique est plus générale et subsume l'approche historique ;

(2) Elle doit être limitée aux systèmes qui sont hiérarchiquement organisés ;

(3) L'intuition selon laquelle il faut distinguer entre fonctionnel/accidentel et dysfonctionnel/non-fonctionnel est erronée, mais on peut en expliquer la source (psychologique) ;

(4) L'approche systémique est plus conforme au naturalisme que l'approche historique, car cette dernière assume l'existence de normes de fonctionnement dont la nature n'est ni causale ni physique[49].

L'amélioration porte sur deux points. D'un côté, elle cherche à éviter ce que Davies appelle la « promiscuité » (ou libéralité) de l'analyse fonctionnelle, c'est-à-dire le fait qu'elle permet d'attribuer des fonctions aux systèmes qui intuitivement n'en ont pas. De l'autre, elle vise à remplacer l'éliminativime par une posture naturaliste et révisionniste[50] où les fonctions ne seraient pas relatives aux intérêts épistémiques de l'observateur mais aux propriétés physiques de l'objet[51].

En ce qui concerne le premier point, Davies (2001, p. 73) limite l'attribution fonctionnelle aux systèmes dotés d'une organisation hiérarchique. Il estime que cette restriction immunise la théorie contre les objections de « promiscuité » formulées à l'encontre de l'analyse fonctionnelle de Cummins. D'après lui, la justification de cette analyse ne dépend pas seulement de sa valeur explicative mais aussi et surtout de la nature du système auquel on l'applique. L'attribution de fonctions ne se justifie que quand des capacités d'un certain niveau sont causées par l'organisation de capacités d'un niveau inférieur. Ces dernières incluent les effets des composants structurels et des interactions causales internes au système. Les êtres vivants, bien entendu, sont des systèmes hiérarchiquement organisés, mais d'autres systèmes inertes le sont également. À vrai dire, reconnaît Davies (2001, p. 84), cette restriction n'évite pas la prolifération des fonctions, car même les objets naturels les plus simples sont constitués de composants de bas niveau (molécules, par exemple) qui donnent lieu à des capacités de plus haut niveau. Cela étant, l'analyse

49 « [T]he historical approach, in attempting to account for the possibility of malfunctions, commits itself to the existence of quite specific norms of performance that are noncausal and nonphysical in nature. » (2001, p. 5)

50 C'est-à-dire qu'il ne cherche pas à analyser ni à rendre compte de l'usage du concept de fonction et remet en cause les intuitions qui s'y rapportent, comme celles qui impliquent une normativité.

51 Un éliminativiste considère qu'il n'y a pas vraiment de différence entre une disposition et une fonction, si ce n'est relativement à un intérêt explicatif. Mais pour Davies, non seulement toutes les dispositions ne sont pas qualifiées pour être des fonctions (car il faut un système hiérarchique), mais l'existence des fonctions n'est pas relative à nos intérêts explicatifs: « *Our explanatory interests may be important in the discovery of systemic functions, but our interests are neither necessary nor sufficient for the existence of such functions.* » (2001, p. 8-9)

fonctionnelle distingue parmi les capacités de bas niveau celles qui contribuent effectivement à une capacité de plus haut niveau de celles qui n'y contribuent pas tout en étant présentes. Les capacités sonores du cœur, par exemple, ne contribuent pas à la circulation sanguine ; et la capacité qu'ont les nuages de mouiller la surface terrestre ne contribue pas, selon Davies, à une capacité d'ordre supérieur du cycle de l'eau, lequel n'est pas un système hiérarchiquement organisé.

En ce qui concerne le second point, Davies défend une conception « substantive » des fonctions systémiques liée à une ontologie naturaliste qu'on pourrait rapprocher de celle de Nagel (1954). Comme lui, Davies (2001, p. 166-167) affirme que son naturalisme est davantage méthodologique qu'ontologique, car il l'entend comme un engagement vis-à-vis de la méthode des sciences naturelles plutôt que vis-à-vis d'une ontologie particulière. Pourtant, chez l'un comme chez l'autre, le naturalisme méthodologique s'appuie sur un certain nombre de présuppositions ontologiques qui limitent le champ et les conditions d'application de la méthode scientifique. Davies, notamment, exclut la normativité du monde naturel. Du début à la fin, il s'attache à démontrer que toutes les tentatives de naturalisation de la normativité biologique sont vouées à l'échec, que les normes de fonctionnement ne font pas partie de la Nature, et que les apparences du contraire sont explicables par la psychologie[52]. D'après lui, la normativité n'est pas dans la Nature mais dans le regard. *A contrario*, les fonctions systémiques existent bel et bien dans la Nature et sont relatives non pas aux intérêts épistémiques de l'observateur, comme chez Cummins, mais aux propriétés physiques du système étudié.

L'engagement ontologique de Davies (2000a, 2001) contre la normativité biologique et contre l'approche étiologique–historique des fonctions repose sur l'idée que les normes de fonctionnement (*norms of performance*) invoqués par les partisans de cette approche sont non causales et non physiques et, par conséquent, contraires au naturalisme. Cet engagement le conduit à distinguer entre les fonctions des systèmes naturels et celles des artefacts. En effet, bien que les artefacts complexes (ordinateurs, caméras) ou plus simples (livres, tasses de thé) aient des fonctions systémiques en vertu de leurs rôles dans divers systèmes de production et de consommation, ces fonctions, dit-il (2001, p. 84 note

52 « I believe we must accept that natural nonengineered traits do not possess the norms of performance that most theorists of functions wish to attribute to them. We must accept this or else admit that we are not naturalists after all. A virtue of the theory defended in this book is its acceptance of this fact and its attempts to explain away in psychological terms the sorts of inclinations that drive us towards the attribution of norms of performance. » (P. S. Davies, 2001, p. 214)

5), sont différentes des fonctions naturelles à cause du rôle central qu'y jouent les intentions et les conventions sociales[53]. Mais l'auteur ne donne pas plus d'explications pour justifier la distinction entre fonctions naturelles et artificielles. Quel est le problème avec ces dernières ? Est-ce parce qu'elles sont plus complexes ou parce qu'elles impliquent une intentionnalité ? Ou encore parce qu'elles sont sujettes à une normativité ? Apparemment, le problème est que les fonctions artificielles sont relatives non pas à ce que l'artefact est capable de faire mais à ce qu'il est censé faire (*supposed to do*). C'est-à-dire qu'elles sont relatives au but ou à la finalité (*purpose*) de l'artefact en question. Or, la détermination de cette finalité n'est pas causale mais mentale.

Étant donné, dit Davies (2009), que la notion de finalité (*purpose*) appliquée au monde naturel dérive de conceptions théologiques du monde que nous ne considérons plus comme vraies et qui ne jouent plus aucun rôle dans les théories scientifiques actuelles, et étant donné que notre tendance à expliquer les êtres vivants en termes téléologiques et à leur appliquer la métaphore du *design* est liée à notre constitution psychologique, il n'y a aucune raison de vouloir conserver et rendre compte de cette notion de finalité en biologie, à l'instar des théories étiologiques-historiques. D'après lui, les fins et les normes n'existent pas dans la nature et elles ne sont pas nécessaires — même métaphoriquement — pour l'expliquer. Cela étant, on ne trouve dans son argumentation aucune opposition à l'usage de ces notions ni à la normativité qui en découle pour rendre compte des actions humaines et des fonctions techniques. On ne peut donc pas répéter à son encontre certaines des critiques formulées à ce propos contre Cummins, mais on peut en formuler d'autres.

Le noyau métallique liquide de la Terre a-t-il une fonction ? Ses mouvements ont la capacité de produire un champ magnétique qui, en déviant les particules ionisantes du vent solaire, contribue à la capacité systémique de notre planète à héberger la vie. Cette capacité de haut niveau du système s'explique par la combinaison de capacités de bas niveau des éléments qui le composent: magnétosphère, couche d'ozone, eau liquide, etc., certaines desquelles sont liées par une organisation complexe. Suivant l'approche de Davies, le noyau métallique terrestre ferait donc un bon candidat à l'attribution d'une fonction systémique… à condition qu'il soit naturel. Car s'il était le produit d'une création intelligente, divine ou extraterrestre, sa fonction prendrait automatiquement

53 D'après lui, il serait naïf de penser que les fonctions des traits naturels — dont le fonctionnement est causal-historique et indépendant de nos intentions ou conventions — puissent servir de modèles pour comprendre celles des artefacts ; et il serait également naïf de penser que ces dernières — dont le fonctionnement dépend de nos intentions et conventions — sont un bon modèle pour comprendre les fonctions naturelles (2001, p. 7-8).

une autre signification et peut-être même aurait-il une autre fonction relative aux intentions de son créateur et, par conséquent, inconnue pour nous. Pourtant, le fait que le noyau terrestre soit un objet naturel ou artificiel ne change strictement rien aux relations causales qu'il entretient avec les autres éléments du système terrestre, et ne change strictement rien pour nous s'il ne nous est pas possible de démontrer ni de réfuter une création intelligente. Si les fonctions systémiques sont des propriétés physiques des systèmes hiérarchiquement organisés, elles devraient donc être les mêmes indépendamment de la nature des systèmes en question et indépendamment de leur origine.

Selon Davies (2001, p. 107), les fonctions systémiques sont des propriétés physiques (des capacités) indépendantes de nous dont la correcte attribution est empiriquement testable. Les fonctions techniques, en revanche, dépendent d'intentions, de conventions et de normes sociales (2001, p. 84 note 5). L'auteur ne précise pas davantage sa conception des artefacts, mais on pourrait peut-être considérer, en reprenant la terminologie de Searle (1995),[54] que les fonctions systémiques sont pour lui des *faits bruts* alors que celles des artefacts seraient plutôt des faits sociaux — ou des *faits institutionnels*. Ces derniers, selon Searle, sont ontologiquement subjectifs (puisque leur existence est relative à certains états mentaux) mais épistémiquement objectifs (car ils ne dépendent pas de préférences individuelles). En ce qui concerne les normes de fonctionnement que l'approche étiologique attribue aux organismes biologiques, Davies rejette leur existence sous prétexte qu'elles ne sont ni causales ni physiques. Ce ne sont, dit-il, que des *illusions*, comme le sont les fins et les fonctions auxquelles fait appel Aristote.[55]

La stratégie de Davies avait plusieurs objectifs, parmi lesquels : éviter l'accusation de promiscuité formulée contre l'analyse fonctionnelle de Cummins, éviter son éliminativisme et justifier une définition pleinement naturaliste du concept de fonction en faisant appel, non pas à ce qu'en pensent les biologistes mais à ce qu'en font les scientifiques en

54 Searle (1995) ne partage ni l'approche systémique ni l'approche étiologique. Il considère plutôt que les fonctions sont relatives à l'observateur et à une attribution préalable de valeurs (voir CHAP. XIII, SECT. 1).

55 À ce propos, Matthew Ratcliffe lui reproche de se placer en porte-à-faux par rapport à l'esprit du naturalisme qu'il revendique : « Where is the warrant for the metaphysical restriction that motivates the account? […] Davies' Humean strategy can be applied to any phenomenon of dubious naturalistic credentials. If it accords with naturalism, then we allow it into the world. If it doesn't, we can always claim that it reflects our own psychological habits rather than the way the world is. Applied across the board, this strategy would render naturalism an irrefutable metaphysical doctrine that the world can be forced to conform to; anything that doesn't accord with it could be attributed to our contingent psychological susceptibilities. » (2003, p. 315).

général, c'est-à-dire en explicitant le rôle qu'il joue dans la recherche scientifique. De notre point de vue, il n'atteint aucun de ces objectifs.

La définition qu'il propose permet d'attribuer des fonctions à n'importe quel système hiérarchiquement organisé pourvu que ses capacités de haut-niveau soient causalement explicables par l'interaction de capacités de plus bas-niveau. C'est-à-dire que l'on pourrait, par exemple, attribuer au champ magnétique terrestre ou au noyau métallique liquide qui le produit une fonction systémique relative à la capacité de notre planète à héberger de la vie. Et si cet exemple n'était pas approprié, il suffirait d'en prendre un autre, car il ne fait aucun doute qu'il existe, selon Davies, de nombreux systèmes naturels non-vivants qui satisfont les conditions d'attribution des fonctions systémiques. Le problème n'est pas que ces attributions soient contre-intuitives ; le problème est qu'elles sont inutiles. Le rôle heuristique que Davies (2001, p. 159-167) veut faire jouer au concept de fonction dans la démarche analytique peut être accompli sans son aide. Preuve en est qu'en dehors de la biologie et des sciences humaines et sociales les autres disciplines scientifiques s'en passent parfaitement. Quel intérêt y a-t-il en effet à attribuer au noyau terrestre une fonction quand on peut se limiter à lui attribuer un rôle causal ou se contenter de dire qu'il contribue à l'une des capacités du système qui nous intéressent ? Si les formulations sont équivalentes, quel avantage y a-t-il à attribuer une fonction à un système naturel abiotique ? Apparemment aucune. Par conséquent, non seulement Davies ne résout pas le problème de la promiscuité de l'analyse fonctionnelle, mais il ne rend pas correctement compte du rôle du concept de fonction dans la recherche scientifique. De plus, on peut considérer que sa définition rejoint l'éliminativisme qu'on attribue à Cummins dans la mesure où elle rend inutile l'attribution de fonctions aux systèmes naturels.

Un dernier commentaire. L'auteur considère qu'une théorie des fonctions n'est naturaliste que (i) si les conditions qu'elle spécifie pour attribuer des fonctions rendent possible un test empirique et (ii) si la théorie elle-même est conforme à l'ontologie et à la méthodologie des meilleures théories scientifiques (2001, p. 108). Or, plusieurs théories peuvent satisfaire (i) et (ii) tout en spécifiant des conditions d'attribution différentes. On pourrait par exemple limiter l'attribution de fonctions systémiques aux systèmes naturels hiérarchiquement organisés et automaintenus, ou bien à ceux capables de se reproduire, ou bien encore à ceux qui sont capables d'une évolution darwinienne. On obtiendrait de cette façon plusieurs théories naturalistes alternatives sans savoir laquelle est correcte ni comment les départager.

Si les fonctions systémiques *existent*, comme semble croire Davies, alors sa définition ne peut pas être seulement stipulative : il s'agit de comprendre *ce que sont* les fonctions. Dans cette perspective, l'auteur exige aux partisans de la conception étiologique qu'ils précisent quels mécanismes causaux sont responsables des fonctions qu'ils attribuent :

> « What natural features of the causal-mechanical processes that constitute a selective history have the power to determine that descendant tokens are for the performance of some task? […] The mere assertion that selected functions are equivalent to selective success is perhaps a plausible opening line, but it cannot be the whole story. As naturalists, we require more than the bare assertion of the view. We require an *account* of the natural mechanisms or the natural causes that give rise to the functional offices and roles. »

En retournant la question à son envoyeur, on s'aperçoit qu'il n'y répond pas non plus. Quels mécanismes causaux distinguent les fonctions systémiques de telle sorte que nous puissions vérifier empiriquement leur existence ? Nous savions déjà que le champ magnétique terrestre nous protège du vent solaire ; nous apprenons maintenant, grâce à Cummins et Davies, que c'est là une de ses fonctions ; mais quelle différence y a-t-il entre ce que nous savions avant et ce que nous savons maintenant ? Aucune. Les mécanismes causaux sont les mêmes. La seule chose qui change est le nom qu'on leur donne.

Pour répondre à la question de Davies depuis sa propre perspective, il faudrait comprendre quels mécanismes causaux sont responsables de l'apparition des fonctions dans un système lorsque l'on passe de systèmes naturels simples à des systèmes complexes hiérarchiquement organisés. Quel est le degré de complexité minimal d'un système doté de fonctions ? Quelles interactions entre capacités de bas-niveau sont-elles nécessaires et suffisantes pour donner lieu à une capacité de haut-niveau justifiant l'attribution de fonctions ? Différents types d'organisations hiérarchiques donnent-ils lieu à différents types de fonctions systémiques ? Voici quelques-unes des questions que devrait soulever une approche naturaliste au sens de Davies.

Ces questions sortent du cadre de la spéculation philosophique pour entrer dans le domaine de la recherche scientifique. Peut-on y répondre ? Il semble que oui. Les récents progrès de la biochimie et de la biologie moléculaire et cellulaire augurent la possibilité de reconstituer en laboratoire le processus évolutif qui a conduit à l'apparition des premiers organismes vivants à partir de systèmes physico-chimiques inanimés. Nous sommes en mesure de fabriquer et d'étudier des organismes synthétiques extrêmement simples situés à la frontière entre le vivant et le non-vivant. Or, si une chose est sûre, c'est que les organismes vivants sont dotés de fonctions. L'étude de ces cellules minimales devrait donc nous apprendre énormément de choses sur les conditions de possibilité des fonctions biologiques et des fonctions systémiques en général. C'est-à-dire que la question des fonctions biologiques semble assez mûre pour que la science vienne apporter une réponse définitive au débat philosophique. Après tout, n'est-ce pas ainsi que la physique est entrée dans le sûr chemin de la science, après n'avoir fait pendant tant de siècles que tâtonner ?

Chapitre III

Approches mixtes

Après avoir examiné dans les chapitres précédents les deux approches principales pour la définition des fonctions biologiques, il nous semble qu'aucune des formulations proposées n'est pleinement satisfaisante. Dans les pages qui suivent, nous allons explorer une voie alternative. Nous allons voir comment l'approche étiologique peut intégrer une analyse fonctionnelle à la Cummins pour compléter l'explication diachronique de la présence d'un item fonctionnel par une explication synchronique compatible avec la théorie darwinienne. Cette solution permet de relativiser le caractère trop fortement historique des fonctions étiologiques en y incluant une dimension organisationnelle qui leur faisait défaut. Elle permet aussi de généraliser l'analyse fonctionnelle de Cummins en l'appliquant à des types, comblant de cette manière l'une des principales lacunes qui lui étaient reprochées.

Nous verrons par la suite, à travers la proposition de Philip Kitcher, comment une approche combinée centrée sur la notion de *design* permet de s'affranchir du mécanisme de la sélection naturelle pour être applicable aussi bien aux artefacts qu'aux êtres vivants indépendamment de leur origine causale. Cette proposition permet d'expliquer la présence d'un trait biologique grâce à l'analyse fonctionnelle des pressions sélectives auxquelles l'organisme porteur est soumis tout en distinguant, par ailleurs, ce pour quoi un trait a été conçu (*design*) de ce pour quoi il a été sélectionné. Étant donné les controverses qui entourent l'idée de *design* en biologie, nous nous attacherons à clarifier ce concept dans le cadre de la proposition de Kitcher puis de manière plus générale en montrant notamment qu'il n'implique ni un créateur ni un dessein intelligent et que, par conséquent, il ne remet pas en question les connaissances scientifiques de la biologie contemporaine.

À la lumière des considérations précédentes, nous envisagerons les attributions fonctionnelles sous la perspective d'une ingénierie inverse appliquée aussi bien aux êtres vivants qu'aux artefacts. Nous reviendrons alors sur plusieurs des idées avancées lors de la discussion des approches

étiologique et systémique, à savoir, d'un côté, l'idée d'une explication valorative non-causale de la présence de l'item fonctionnel, et, d'un autre côté, l'idée d'une analyse fonctionnelle normative applicable à des types et pas seulement à des items particuliers (voir nos commentaires à propos de la conception de Cummins).

1. Complémentarité des approches étiologique et systémique

Les diverses tentatives de naturalisation des fonctions biologiques sont complémentaires dans leurs différences. Elles sont complémentaires tout d'abord quant à leur *perspective temporelle*. Certaines adoptent une perspective tournée vers le passé – la fonction d'un trait étant relative à son histoire –, d'autres se tournent plutôt vers le futur, et d'autres encore s'en tiennent au présent ou offrent une définition intemporelle.

De plus, elles répondent à des *questions différentes*. En effet, les théories étiologiques rendent compte de la présence d'un trait donné dans l'organisme, tandis que les approches systémiques visent à expliquer la contribution de ce trait à un but ou à une capacité de l'organisme ou de l'un de ses systèmes.

À ces explications correspondent des *relations causales différentes*. Les théories étiologiques invoquent en effet des causes distales, externes à l'organisme, car la fonction d'un trait y est relative à son histoire, et notamment à celle de ses ancêtres, tandis que les théories systémiques définissent au contraire les fonctions et la téléologie en termes de causes proximales internes à un système ou à un sous-système de l'organisme.

Par ailleurs, les théories étiologiques partent de l'intuition, chez Wright, qu'*un trait fonctionnel contribue à sa propre présence* à travers ses effets, mais paradoxalement les définitions de Neander et de Millikan ne s'intéressent pratiquement ni aux effets ni à leur contribution. Pour elles, peu importe le mécanisme ou la manière dont un trait contribue à sa propre présence, car seul compte le fait avéré de sa sélection. Peu importe la contribution actuelle puisque les traits vestigiaux n'en sont pas moins dotés de fonctions, et peu importent aussi les effets actuels puisque les organes et les membres des doubles accidentels sont au contraire dénués de fonctions bien qu'ils soient structurellement identiques et possèdent exactement les mêmes effets que leurs homologues « naturels ».

En revanche, les théories systémiques s'intéressent aux effets ou aux conséquences (notamment sous forme de dispositions causales) de la présence d'un trait dans l'organisme, mais elles négligent la question du rapport entre les effets et la présence. Elles partent de l'intuition que *la fonction d'un trait est directement liée à son rôle dans un système*, mais elles ne s'interrogent pas sur les causes ni les raisons de sa présence dans ce système. C'est comme si, pour elles, la contribution du cœur au système

circulatoire des vertébrés n'avait rien à voir avec le fait que les vertébrés ont effectivement un cœur.

Ici, la complémentarité des approches met en évidence leurs lacunes respectives, de sorte que l'on peut se demander s'il ne serait pas plus opportun de chercher à les concilier plutôt que de les opposer. Une conciliation pouvant prendre au moins deux formes : le pluralisme, qui consiste à dire qu'elles sont également acceptables dans la mesure où elles correspondent à des usages différents du concept de fonction, et l'unification, qui consiste à rassembler dans une conception unique les intuitions et les outils conceptuels qui font la force des autres.

Ainsi, pour répondre à la question du rapport entre la présence d'un trait fonctionnel et son rôle dans l'organisme, nous devons nous tourner soit vers les formulations propensionnistes de l'approche étiologique, comme celles de Ruse, de Bigelow et Pargetter, ou encore celle de Walsh, soit vers des approches ouvertement unificatrices comme celles de Griffiths et de Kitcher. Les premières ne se contentent pas de dire qu'un trait a une fonction parce que certains de ses effets ont été sélectionnés ; elles ajoutent à cela que s'ils ont été sélectionnés, c'est parce qu'ils contribuent d'une manière ou d'une autre à la *fitness* des organismes porteurs. C'est-à-dire qu'elles ne se contentent pas du fait avéré de la sélection, mais lui cherchent une justification en termes de contribution au fonctionnement de l'organisme (dans un environnement donné), sans toutefois entrer dans les détails de cette contribution. Les secondes vont plus loin en employant l'analyse fonctionnelle de Cummins pour rendre compte de ces détails. Elles combinent l'intuition systémique selon laquelle la fonction d'un trait est le rôle qu'il joue dans un système, entendu en termes de contribution aux capacités de survie et de reproduction d'un organisme, et l'intuition étiologique selon laquelle la fonction d'un trait est ce qui en explique la présence, par le biais du mécanisme de sélection naturelle. Ainsi, ces formulations alternatives permettent d'expliquer la présence d'un trait à partir de son rôle dans l'organisme, sans affirmer pour autant que le rôle soit la cause de la présence.

2. Pluralisme et unification

L'existence persistante de deux approches concurrentes pour la définition du concept de fonction a conduit certains auteurs à admettre une sorte de pluralisme consistant à dire que, bien que l'approche étiologique rende compte de la plupart des usages du concept de fonction en biologie évolutive, il reste de la place pour une approche systémique qui couvre les usages de la biologie fonctionnelle (Allen & Bekoff, 1995b; Amundson & Lauder, 1994; Bouchard, 2013; Brandon, 2013; Caponi, 2001b; Godfrey-Smith, 1993; Millikan, 1989a, 2002; S. D. Mitchell, 1993;

Perlman, 2009; Preston, 1998; Sober, 1993). À propos des différentes formes de pluralisme, voir Garson (2016).

D'après Millikan, non seulement les fonctions propres et les fonctions à la Cummins correspondent, en biologie, à des phénomènes différents — puisqu'un trait organique peut avoir l'une et pas l'autre —, mais elles sont elles-mêmes dans une certaine mesure indéterminées, de sorte que l'on doit éviter de pousser trop loin la précision dans leur définition et abandonner l'idée qu'il existerait un usage « correct » de ces termes :

> « Do not attempt to give these notions entirely clean boundaries. Nature has many important joints, but these joints are seldom clean. Definitions that cut sharp edges where there are none in nature are of little use in the understanding of nature. […] But most important, to be avoided at all costs is the attitude that there must be some 'correct' way of using these terms, some preordained way waiting to be discovered. » (Millikan, 2002, p. 122)

Considérant la capacité de survie ou de reproduction d'un organisme, on devrait pouvoir selon Millikan effectuer une analyse fonctionnelle à la Cummins pour déterminer quels traits organiques ou comportementaux ont une « utilité vitale ». Cette analyse pose un certain nombre de problèmes, notamment par le fait qu'il ne s'agit pas d'analyser un individu particulier mais un type d'organismes et aussi par le fait que la survie et la reproduction dépendent de conditions environnementales particulières. Nonobstant, les biofonctions systémiques attribuées à un trait dans le cadre d'une analyse fonctionnelle pertinente de l'organisme porteur correspondent, selon Millikan (2002, p. 139), à ses fonctions propres. La seule exception à cette règle concerne les « exaptations » (S. J. Gould & Vrba, 1982) qui, d'après elle, sont des biofonctions systémiques sans être des fonctions propres. Interprétées de cette façon, les biofonctions systémiques sont-elles plus générales et englobent-elles les fonctions propres ou s'agit-il de deux types différents de fonctions qui coïncident dans la plupart des cas ? L'une des thèses défendues par Davies (2000b, 2001) était justement que l'approche systémique est plus générale et, de fait, subsume l'approche étiologique, mais ses conclusions sont bien évidemment à l'opposé de celles de Millikan.

D'autres auteurs (Buller, 1998; Griffiths, 1993; Kitcher, 1993; Schlosser, 1998, 2003; Walsh & Ariew, 1996) ont proposé de combiner les approches étiologique et systémique dans une conception unique tenant compte des intuitions propres à chacune.

La proposition de Griffiths est d'incorporer l'analyse fonctionnelle de Cummins dans une conception de type étiologique tournée vers le passé : les fonctions propres d'un trait seraient ainsi celles attribuées par une analyse fonctionnelle des capacités de survie et de reproduction (*fitness*) des porteurs de ce trait (Griffiths, 1993, p. 412). En d'autres

termes, les fonctions propres d'un trait chez un organisme actuel sont les effets de ce trait ayant contribué à la *fitness* de ses ancêtres. Ce sont les effets en vertu desquels il a été sélectionné. Cela veut dire, ajoute Griffiths, qu'un trait n'aura de fonction propre que s'il constitue une *adaptation* pour cette fonction. Sa proposition est la suivante :

> « Where i is a trait of systems of type S, a proper function of i in S's is F iff a selective explanation of the current non-zero proportion of S's with i must cite F as a component in the fitness conferred by i. » (Griffiths, 1993, p. 415)

Cette définition, selon lui, rend possibles deux explications. En premier lieu, la *fitness* d'un type d'organisme serait explicable à partir d'une analyse fonctionnelle à la Cummins. En second lieu, la présence des traits fonctionnels serait explicable en vertu de leur contribution à cette *fitness* et par le biais du mécanisme de sélection naturelle.

De ce point de vue, la proposition de Griffiths est similaire à celles — postérieures — de Godfrey-Smith (1994) et de Walsh (1996). Nous avons vu en effet que la théorie relationnelle de ce dernier offrait également la possibilité d'appliquer à un même trait deux explications différentes : l'explication étiologique de la présence ou de la prévalence d'un trait dans une population, et l'explication anhistorique de la *fitness* d'un individu ou d'un groupe d'individus en vertu de la contribution de ce trait à cette *fitness*.

L'idée d'incorporer l'analyse fonctionnelle au sein de l'approche étiologique répond au besoin d'expliciter les raisons pour lesquelles un trait organique a été historiquement sélectionné. Ces raisons s'expriment généralement en termes de valeur de survie, de valeur reproductive ou de valeur adaptative (voir p.78) : si un trait donné s'est diffusé au sein de la population et a perduré dans le temps, c'est sans doute parce qu'il améliorait ou contribuait favorablement aux capacités de survie et de reproduction des individus qui en étaient dotés — comparativement à celles des individus qui en étaient dépourvus, dans un contexte environnemental déterminé, etc.

À vrai dire, le type d'analyse que décrit Cummins pourrait très bien correspondre au raisonnement des biologistes évolutionnistes et des partisans de l'approche étiologique lorsqu'ils s'interrogent sur la présence d'un trait. Résumé en deux étapes, le raisonnement serait le suivant :

i. Ce trait contribue-t-il — et comment — aux capacités de survie et de reproduction des organismes qui le portent ?

ii. Ce trait a-t-il été sélectionné à cause, en vertu ou pour sa contribution à la survie ou au succès reproductif des organismes qui le portent ?

La première étape correspond à l'analyse fonctionnelle de Cummins si ce n'est qu'elle porte sur des types d'organismes plutôt que sur des tokens, ce qui n'est pas incompatible avec la démarche de l'auteur[56]. De plus, la capacité analysée est prédéterminée et pas arbitrairement choisie par l'observateur, mais comme dit Ruth Millikan (2002, p. 132) il semble raisonnable de penser que la capacité de l'organisme de s'automaintenir et de se reproduire est justement la capacité implicitement analysée quand on attribue des biofonctions systémiques à la Cummins. Cette première étape ne requiert pas la prise en considération de l'origine ni de l'histoire causale de ce trait ou du type d'organisme analysé. Elle correspond par exemple à l'identification de la fonction physiologique du cœur par William Harvey indépendamment de la théorie darwinienne.

La seconde étape correspond à l'étude de l'apparition et de l'évolution de ce trait au cours de l'histoire, à l'étude de ses conséquences, de sa transmission, etc. C'est sur la base de ce genre d'analyse que l'on doit pouvoir déterminer si le trait ou ses conséquences sont une adaptation, une exaptation ou un *spandrel*. C'est aussi sur la base de ce genre d'analyse que l'on peut distinguer entre les conséquences accidentelles et fonctionnelles d'un trait comme le cœur, sélectionné pour sa capacité à pomper le sang plutôt que pour faire du bruit. Mais l'attribution de fonctions ou de fonctions propres sur la base de la sélection naturelle est plus difficile que l'attribution de fonctions systémiques, et il est raisonnable de penser que celle-ci est utile à celle-là.

Appliquons ce raisonnement aux doubles accidentels qui servent de contre-exemple à la formulation historique de l'approche étiologique (voir p. 66). En réponse à la première question, nous tirerions immédiatement la conclusion que les protubérances latérales des pseudo-lions contribuent à leur capacité de survie et de reproduction dans la mesure où elles leurs permettent de se déplacer avec une vitesse remarquable, de se défendre efficacement et d'attraper des proies. A défaut de leur trouver une autre ou une meilleure fonction systémique, celle-ci constitue une bonne hypothèse de départ pour expliquer la présence de ces membres. S'il s'avère que les pseudo-lions sont effectivement le produit d'un accident cosmique et pas d'un processus sélectif, alors l'explication de leur présence et de leur configuration se limitera à invoquer le hasard, mais

56 De fait, les schémas fonctionnels que Cummins cite en exemple et qui peuvent correspondre aussi bien à un circuit imprimé qu'à un programme informatique ou une machine-outil ne sont pas des artefacts concrets (des réalisations matérielles) mais leur idéalisation (voir note 44, p. 106). L'analyse fonctionnelle peut donc aussi bien porter sur un engin concret que sur ses plans de construction. Enç (1979, p. 347) donne à propos de la découverte de la fonction du cœur par Harvey un bon exemple d'analyse fonctionnelle appliquée non pas à un organe concret mais à un type. L'interprétation de la proposition de Cummins en termes de types est suggérée au passage par Neander (2010, p. 102).

cela ne nous empêche pas de leur attribuer une fonction systémique relative à la capacité de survie et de reproduction de ces organismes, ni de les classer dans la catégorie fonctionnelle qui leur correspond[57]. En revanche, s'il s'avérait que ces créatures procèdent d'une Terre jumelle, l'hypothèse serait susceptible d'être validée ou infirmée en allant sur cette planète.

L'idée que les membres des pseudo-lions pourraient être des ailes malformées ou des structures vestigiales, comme dit Karen Neander, est dénuée de toute justification apparente. D'autant plus que la définition historique de cette auteure rend difficile la distinction entre traits actuellement fonctionnels et traits vestigiaux : tous deux ont droit ou devraient avoir droit à l'attribution d'une fonction propre dans la mesure où elle dépend d'un processus sélectif lié à leur contribution — *passée* — à la fitness des *ancêtres* des organismes actuels.[58] En effet, dans une approche de type historique comme celle de Neander, un trait doit être considéré comme fonctionnel même s'il ne contribue pas *actuellement* à la *fitness* des organismes qui le portent. Le contraire aurait d'ailleurs été difficile à établir : peut-on dire que le cœur des êtres humains contribue actuellement à leur *fitness* s'il n'y a pas d'alternative à cet organe ni de pression sélective ? Ce qui est sûr, en revanche, c'est que le cœur contribue à notre capacité de survie et de reproduction tandis qu'une structure vestigiale comme le coccyx n'y contribue plus depuis longtemps ; le premier conserve donc sa fonction systémique alors que le second l'a perdue. Or, n'est-ce pas là que réside la différence principale entre un trait fonctionnel et un trait vestigial ? Si c'est le cas, alors la reconnaissance d'une structure vestigiale requiert une analyse fonctionnelle à la Cummins.

L'intégration d'une telle analyse dans une théorie de type étiologique n'est pas la seule manière d'unifier les principales approches sur la question. En partant d'une conception systémique et holistique des êtres vivants, l'approche organisationnelle (W. D. Christensen, 1996; W. D. Christensen & Bickhard, 2002; McLaughlin, 2001; Mossio et al., 2009; Schlosser, 1998) cherche à rendre compte de la téléologie et la normativité et à expliquer la présence des traits biologiques à partir de l'analyse des relations causales internes aux organismes. Nous y reviendrons au CHAP. XIV.

57 Bien qu'elles n'aient pas été *créées pour* quoi que ce soit, ces protubérances sont nonobstant *utilisées comme* des pattes par les pseudo-lions, et si on ajoute à cela qu'elles sont structurellement similaires à celles d'autres animaux et que, par-dessus le marché, on peut les voir *comme si* elles étaient le fruit d'un *design*, alors il semble raisonnable de les considérer comme des pattes. Dans le cas contraire, il faudrait reconsidérer la catégorie fonctionnelle à laquelle appartiennent de nombreux artefacts humains qui n'ont pas été créés pour l'usage qui est actuellement le leur.

58 Kitcher (1993, p. 264) discute de l'ambiguïté relative à la temporalité des processus de sélection dans les attributions fonctionnelles étiologiques.

3. Formulation en termes de *design*

D'après George C. Williams (1966, p. 209), le *design* est une condition nécessaire et suffisante pour avoir une fonction. Selon Ruth Millikan (1984, p. 17), avoir une fonction propre c'est avoir été « conçu pour » (*designed to*) réaliser une certaine fonction. Et Michael Ruse (2002) défend la métaphore du *design* appliquée aux organismes et à leurs parties pour rendre compte de la téléologie et des fonctions biologiques. Dans le cadre de la théorie darwinienne, il s'agit pour ces auteurs de penser le *design* sans un *designer* ; une horloge sans horloger.

Philip Kitcher (1993) va plus loin en cherchant à unifier l'approche étiologique avec l'analyse systémique de Cummins autour de la notion de *design* tout en reconnaissant le pluralisme des usages biologiques du concept de fonction[59]. Selon cet auteur, on peut rendre compte de ces usages en considérant que la fonction d'un item *X* est ce pour quoi *X* a été conçu (*what X is designed to do*), mais pas forcément ce pour quoi *X* a été sélectionné. Bien que la source du *design* d'un organisme se trouve dans les opérations de la sélection naturelle — de même que la source du *design* d'un artefact se trouve dans les intentions de son créateur — le lien entre les deux n'est pas toujours direct. En effet, la sélection naturelle n'est pas toujours une explication suffisante de la présence d'un item

59 Les différentes interprétations de la fonction de *X* correspondent à différents modes de sélection de *X*. La source d'un *design* peut en effet se trouver soit dans l'action d'un agent intentionnel, soit dans l'action de la sélection naturelle, laquelle peut agir de plusieurs façons et à des échelles de temps différentes. Kitcher souligne une double ambiguïté de l'attribution de fonctions sur la base de la sélection naturelle : une ambiguïté temporelle et une ambiguïté dans la compétition sur laquelle s'appuient les processus de sélection. Pour qu'il y ait fonction, faut-il que la sélection explique seulement l'apparition du trait ou à la fois son apparition et son maintien, ou seulement son maintien ? Et si l'on considère seulement le maintien, doit-on considérer le passé lointain, le passé récent ou le présent? Par ailleurs, quelles sont les alternatives aux entités biologiques dont la présence s'explique par la sélection ? Et dans quelle mesure la sélection est-elle une explication complète de la présence de cette entité ? À chacune de ces questions correspond une interprétation potentiellement différente du concept de fonction répondant à des méthodologies et à des projets explicatifs différents :

(a) les conceptions historiques du concept de fonction expliquent soit la diffusion initiale d'un trait (histoire lointaine) soit sa présence actuelle (histoire moderne),

(b) les conceptions propensionnistes (comme celle de Bigelow et Pargetter) expliquent sa présence dans un futur plus ou moins proche, et

(c) l'analyse fonctionnelle explique sa contribution au fonctionnement de l'organisme dans son ensemble et rend compte de la manière dont celui-ci répond aux pressions environnementales.

— ou plutôt : le fait que *X* soit présent ne s'explique pas seulement par l'action de la sélection naturelle.[60] La théorie de Kitcher permet par conséquent d'attribuer une fonction à un item indépendamment du fait qu'il ait ou pas été *sélectionné pour* cette fonction.[61]

L'une des particularités de la proposition de Kitcher est le fait de considérer la sélection naturelle depuis une perspective centrée sur l'environnement : étant donné un contexte environnemental dont les caractéristiques sont plus ou moins fixées, on peut s'interroger sur les pressions sélectives auxquelles sont soumis les membres d'un groupe d'organismes et analyser leurs traits (leur *design*) à la lumière de ces pressions (1993, p. 261). Par exemple, un mammifère herbivore placé dans un environnement dont les végétaux sont pourvus d'épaisses parois de cellulose devra lui-même disposer d'un moyen d'en venir à bout ; il sera donc soumis à une pression sélective pouvant le conduire à développer une réponse appropriée à son environnement, que ce soit sous la forme de fortes molaires ou de bactéries intestinales spécialisées. Selon Kitcher (1993, p. 270), la tendance à attribuer une fonction à un item *X* (les molaires) peut procéder de la reconnaissance du fait que *X* est une réponse à une pression sélective de l'environnement.

> It is enough [for the attribution of functions] that genuine demands on the organism have been identified and that the entities to which [we] attribute functions make causal contributions to the satisfaction of those demands. (Kitcher, 1993, p. 271)

Un exemple récent analogue de celui proposé par Kitcher est fourni par une population de « lézards des ruines » (*Podarsis sicula*) introduite artificiellement en 1971 sur la petite île Croate de Pod Mrčaru dans l'Adriatique (Herrel et al., 2008). Après seulement 36 ans d'isolement, soit environ 30 générations, les lézards se sont adaptés à un régime alimentaire composé principalement de plantes riches en cellulose alors qu'ils sont normalement insectivores. Leurs têtes sont devenues plus

60 Par exemple, la chance peut favoriser la survie d'une entité dont la *fitness* (au sens de valeur adaptative) est inférieure à celle de la plupart de ses rivales dans la mesure où le hasard provoque leur élimination (Kitcher, 1993, p. 269-270).

61 De ce point de vue, la proposition de Kitcher est comparable à la version étiologique faible de Buller (1998) et relativement proche de celles de Godfrey-Smith (1993) et de Walsh (1996). Ce dernier auteur va plus loin en affirmant que la sélection naturelle n'est pas la cause des adaptations biologiques et que, par conséquent, le programme de réduction (causale) de la téléologie ne peut pas réussir (Walsh, 2000).

grandes, leurs mâchoires plus puissantes et leur morphologie intestinale a profondément changé avec l'apparition de valves cæcales similaires à celles de certains herbivores mais rarement observées chez des lézards et jamais chez les populations de cette espèce. Les valves cæcales ralentissent le passage de la nourriture et constituent des chambres de fermentation où des microorganismes hôtes peuvent digérer la cellulose des plantes. De fait, ont trouve dans leurs intestins des microorganismes absents chez les autres populations de cette espèce. La proposition de Kitcher permet d'attribuer des fonctions aux modifications morphologiques des lézards des ruines dans la mesure où elles constituent des réponses aux pressions sélectives (régime alimentaire) de leur nouvel environnement. Et cette attribution est indépendante de l'identification des mécanismes évolutifs impliqués.[62]

Une autre particularité de la proposition de Kitcher est la distinction subtile qu'il établit entre les sources — directe et indirecte — du *design.* Une distinction applicable aussi bien aux artefacts qu'à la biologie pré-darwinienne et à la biologie évolutionniste. Avant Darwin, l'attribution de fonctions aux créatures et à leurs parties pouvait correspondre à deux conceptions différentes de l'œuvre du Créateur : ou bien celui-ci avait conçu dans un but précis jusqu'aux plus infimes détails de sa création, ou bien il avait réalisé cette dernière par le biais de causes secondes, les organismes étant équipés de manière à répondre à leurs besoins sans que les détails de ces réponses fussent spécifiées à l'avance (1993, p. 260). De manière analogue, dans le domaine des artefacts, les parties d'une machine peuvent avoir des fonctions non prévues par leur créateur, car celui-ci ne connaît pas forcément tous les détails des conditions de fonctionnement de sa création. Les fonctions en question sont déterminées par la contribution qu'elles apportent au fonctionnement de la machine dans son ensemble et par le fait qu'elles répondent — indirectement — aux intentions explicites de son créateur. Plus précisément, lorsque le fonctionnement de la machine répond aux intentions explicites de son créateur, les fonctions des parties de la machine sont déterminées par leur contribution au fonctionnement de celle-ci (1993, p. 260). Dans le domaine de la biologie moderne, où la source du *design* des organismes n'est pas à chercher du côté d'un créateur intelligent mais du côté de la sélection naturelle, on peut selon Kitcher établir une distinction similaire entre la sélection directe de tous les détails du *design* des êtres vivants et la sélection de ceux-ci par le biais de « causes secondes ».

62 Bien que la sélection naturelle y ait sans aucun doute joué un rôle essentiel, le caractère fonctionnel des valves cæcales est compatible avec l'intervention d'autres mécanismes causaux, y compris dans un cadre pré-darwinien. C'est-à-dire qu'on peut connaître la fonction de cette structure sans connaître les causes de sa présence.

Cette distinction a une incidence importante sur la valeur explicative des attributions fonctionnelles : Si l'item *X* est une réponse (ou contribue à répondre) à une pression sélective de l'environnement mais n'est pas — ou pas seulement — un produit direct de la sélection naturelle, alors la présence de *X* n'est pas entièrement explicable en termes de sélection[63] et la fonction qu'on lui prête n'est donc pas toujours une explication suffisante de sa présence. Le lien entre fonction et sélection n'est certes pas rompu, et Kitcher insiste sur le fait que cette dernière est la source du *design* des êtres vivants, mais son rôle n'est pas aussi direct ni aussi clair que dans les versions dites « fortes » de la conception étiologique[64]. Un item n'est pas fonctionnel parce qu'il a été sélectionné (en vertu de sa contribution à la *fitness* des ancêtres des porteurs de ce trait) mais parce qu'il a été conçu pour (*designed to*) répondre à une pression sélective de l'environnement — où la sélection naturelle constitue la source mais pas toujours la cause du *design* en question. Sa fonction est déterminée, à la manière de Cummins, par sa contribution au fonctionnement de l'organisme dans son ensemble (ou à l'un de ses sous-ensembles) en réponse à une pression sélective (Kitcher, 1993, p. 262).

On pourrait s'aventurer à interpréter la position de Kitcher en disant que si la sélection naturelle ne fournit pas toujours une *explication* suffisante de la présence d'un item fonctionnel, on peut néanmoins *comprendre* sa présence *dans le cadre* de la sélection naturelle — ou plus précisément dans le cadre de la théorie synthétique de l'évolution. C'est-à-dire que l'on peut comprendre les raisons de la présence d'un item fonctionnel sans en connaître précisément les causes et sans que ces causes soient exclusivement dues au mécanisme de sélection naturelle proprement dite (entendue de manière simpliste comme la « survie du plus apte »). On peut comprendre qu'un item soit présent parce qu'il répond à une pression sélective, comme les valves cæcales des lézards de Pod Mrčâru, sans que la sélection naturelle en soit directement et entièrement responsable. En effet, d'autres mécanismes causaux non sélectifs peuvent également être impliqués, parmi lesquels des mécanismes aléatoires comme la dérive génétique. On comprend la présence d'un trait

63 Pour pouvoir parler de sélection, il faut qu'il y ait plusieurs alternatives parmi lesquelles une sélection puisse opérer. Or, les alternatives concurrentes du trait en question sont parfois peu ou pas connues. De plus, la présence d'un trait n'est pas toujours seulement le produit d'une sélection proprement dite parmi des alternatives rivales, c'est-à-dire une sélection où « le meilleur gagne », mais aussi le fruit d'un hasard susceptible d'éliminer les « meilleurs » (voir note 50, p. 110), ou encore le fruit d'une « dérive génétique » (*genetic drift*). On trouve un équivalent dans le domaine de la technologie avec la guerre des formats vidéo dans les années 1970 et '80 où le VHS s'est imposé face à un Betamax technologiquement supérieur.

64 Cf. la distinction établie par Buller (1998) entre les versions forte et faible.

fonctionnel dans la mesure où, sans lui, toutes choses étant égales par ailleurs, l'organisme serait incapable de répondre à une pression sélective donnée et pourrait par conséquent être conduit à disparaître. En revanche, le fait que l'organisme doive répondre à cette pression (pour survivre) ne détermine pas la forme de sa réponse. C'est-à-dire qu'on ne comprend pas nécessairement pourquoi un item particulier (comme les valves cæcales) est présent plutôt qu'un autre, mais on comprend pourquoi l'organisme possède un item (quel qu'il soit) répondant à une pression sélective donnée.

Selon cette interprétation, attribuer aux valves cæcales la fonction de favoriser la digestion de la cellulose des plantes, ce n'est pas expliquer pourquoi les valves cæcales existent mais pourquoi il existe *un trait X* (quel qu'il soit) capable de le faire. Plus précisément, attribuer une fonction aux valves cæcales c'est expliquer :

i. *pourquoi* l'organisme ou l'une de ses parties est conçu pour (pourquoi il a développé la capacité de) s'alimenter à base de plantes riches en cellulose (dans un environnement approprié) et

ii. *comment* les valves cæcales contribuent à la réalisation de cette fin (ou de cette capacité).[65]

La seconde partie de l'explication est manifestement causale — et correspond à une analyse fonctionnelle à la Cummins —, mais la première ne l'est pas. C'est une explication étiologique — à la manière de Wright — qui s'inscrit dans le cadre de la théorie de l'évolution mais qui ne se limite pas à la sélection naturelle ni ne précise le mécanisme causal responsable de la présence de l'item *X*. Ce n'est pas parce que l'organisme répond à une pression sélective qu'il survit et se reproduit, mais s'il ne le faisait pas il ne survivrait pas. Donc, s'il s'est perpétué, c'est forcément (en partie) parce qu'il y a répondu : le fait d'y répondre est nécessaire mais pas suffisant pour expliquer la perpétuation de l'organisme.

Une des difficultés de la proposition de Kitcher est le sens qu'il convient de donner à l'expression « conçu pour » (*designed to*) qu'il ne prend pas la peine de définir. On pourrait penser au premier abord qu'elle correspond à la notion d'adaptation autour de laquelle Michael Ruse élabore sa propre définition, mais une lecture plus attentive montre que Kitcher ne s'appuie pas sur cette notion et que sa proposition, contrairement à celle de Ruse, excède le cadre de la théorie de l'évolution ainsi que le domaine des êtres vivants.[66] Il n'emploie pas non plus la

65 Comparer avec la décomposition du raisonnement étiologique présentée plus haut, p. 121.

66 Cela étant, dans le domaine de la biologie, un trait qui répond aux pressions sélectives de l'environnement est-il autre chose qu'un trait adaptatif — sans être nécessairement une adaptation ? Sober (1984, p. 208) définit une adap-

notion de *fitness*, ce qui l'éloigne d'autant plus de la proposition de Walsh (voir p. 71). La difficulté vient principalement du refus de Kitcher d'établir un lien direct entre *design* et sélection tout en insistant sur le fait que la source du *design* en biologie est la sélection naturelle et qu'un trait fonctionnel est une réponse à une pression sélective. De plus, il affirme que sa proposition est compatible — dans le cadre d'un contexte environnemental stable — aussi bien avec les projets explicatifs visant à rendre compte du maintien d'un trait dans le passé (de préférence récent) ou au présent, qu'avec les projets prédictifs comme celui de Bigelow et Pargetter portant sur la conservation et/ou la diffusion du trait dans un futur proche. Un trait fonctionnel, s'il a été *conçu pour*, ne peut pas être une réponse *accidentelle* à une pression sélective de l'environnement, mais ce pour quoi il a été conçu (sa fonction actuelle) n'est pas forcément ce pour quoi il est apparu (dans un passé plus lointain) ni ce pour quoi il a éventuellement été « façonné » par la sélection naturelle. Selon Godfrey-Smith (1994, p. 356), les fonctions de Kitcher seraient « des dispositions et des forces » (*powers*) qui expliquent le maintien récent d'un trait dans un contexte sélectif. » Or, ce qui fait qu'un trait ait une disposition c'est sa structure ou son organisation. Il convient donc peut-être d'interpréter l'expression « conçu pour » en termes organisationnels plutôt qu'en termes sélectifs.

Une chose est claire : si *X* a été conçu pour *F*, alors *F* ne peut pas être une conséquence accidentelle de la présence de *X*, et *X* ne peut pas non plus être là par accident. Ce qui caractérise un *design* c'est justement le fait que sa constitution et ses effets ne *semblent* pas être dus au hasard mais au contraire à une planification intelligente[67]. Et peu importe la nature du processus (intention consciente, sélection naturelle, etc.) si le résultat est le même, si tout semble être à sa place, à la place qu'il faut pour produire les conséquences qu'il faut ; par exemple, pour répondre à une pression sélective. Cela étant, l'idée que la présence et que les conséquences de *X* ne puissent pas être dues au hasard ne doit pas être interprétée en termes causaux car, de fait, la cause de la présence de *X*

---

tation en termes de *sélection pour* historiquement avérée : « *A is an adaptation for task T in population P if and only if A became prevalent in P because there was selection for A, where the selective advantage of A was due to the fact that A helped perform task T.* » [cité par Tim Lewens (2007, p. 2)] Or un trait adaptatif, entendu comme un trait qui contribue à la *fitness*, n'est pas forcément une adaptation s'il n'a pas l'histoire qu'il faut.

67 C'est le point sur lequel tombent d'accord aussi bien William Paley que Richard Dawkins (bien que les conclusions qu'ils en tirent ne soient évidemment pas les mêmes) : « *We may say that a living body or organ is well designed if it has attributes that an intelligent and knowledgeable engineer might have built in order to achieve some sensible purpose, such as flying, swimming, seeing, eating, reproducing, or more generally promoting the survival and reproduction of the organism's genes* » (Dawkins, 1986, p. 21)

peut être attribuée au hasard. Pour le comprendre, il faut se reporter à l'exemple de la machine que donne Kitcher (1993, p. 260) : Un ingénieur conçoit une machine pour répondre à un certain besoin, mais il ne se rend pas compte que pour qu'elle fonctionne correctement il doit établir une jonction entre deux de ses parties ; heureusement, tandis qu'il travaille à sa construction, l'ingénieur laisse involontairement tomber une petite vis qui se loge précisément entre les deux parties en question, établissant la jonction nécessaire. La présence de la vis dans la machine est un produit du hasard ; nonobstant, elle se trouve à sa place, exactement à la place qu'il faut pour que la machine fonctionne correctement. D'un point de vue causal, la présence de la vis et les conséquences de sa présence sont purement accidentelles. Mais du point de vue fonctionnel, c'est-à-dire du point de vue du *design* de la machine, la présence et les conséquences de la vis semblent absolument nécessaires.

Larry Wright (1973, p. 152) utilise un exemple similaire pour en tirer une conclusion contraire : si un objet tombe par accident dans une machine et contribue à son bon fonctionnement, nous ne lui attribuerons pas la fonction correspondante. La contradiction entre les deux auteurs n'est qu'apparente, car leurs conclusions correspondent à deux perspectives différentes du même problème. Wright veut montrer, contre Canfield, qu'une contribution utile n'est pas suffisante pour déterminer une fonction ; en mettant en avant le caractère accidentel de la présence de l'objet dans la machine il montre que celui-ci n'a pas de raison d'être. Kitcher montre quant à lui que si la présence d'une jonction à un endroit donné de la machine est nécessaire à son bon fonctionnement, alors l'objet qui réalise cette jonction a bel et bien une raison d'être. Les deux postures sont compatibles si l'on considère que Wright fait référence à l'objet concret (*token*) et aux raisons circonstancielles de sa présence tandis que Kitcher fait référence à l'objet en général, c'est-à-dire au type auquel il appartient, et aux raisons structurelles ou fonctionnelles de sa présence.

4. *Design* et ingénierie inverse

4.1. Fonctions, raisons d'être et intentions

L'idée que « la fonction d'un item X est ce pour quoi X a été conçu », que ce soit en biologie ou à propos des artefacts, doit être interprétée à la manière d'un ingénieur en train d'analyser le fonctionnement et l'organisation d'une machine. Lorsqu'il observe la petite vis de l'exemple précédent, il n'a *à priori* aucune raison de penser qu'elle se trouve là par hasard. Au contraire, dans la mesure où elle joue un rôle crucial pour le fonctionnement global de l'appareil, il a raison de croire

que sa présence à cet endroit est intentionnelle, car si elle ne s'y trouvait pas, il faudrait l'y mettre. C'est-à-dire qu'en analysant le fonctionnement et la structure de la machine, l'ingénieur associe le *design* de celle-ci aux intentions présumées de son créateur. Contrairement à l'analyse fonctionnelle de Cummins qui s'applique à n'importe quelle capacité systémique choisie par l'observateur, l'ingénierie inverse que nous sommes en train de décrire cherche à découvrir quelles sont les capacités du système qui sont pertinentes pour comprendre l'organisation de ses parties. En d'autres termes, il ne s'agit pas seulement de comprendre quel rôle joue un élément dans la réalisation d'une capacité donnée du système mais aussi et surtout de savoir quelle est la capacité du système pour laquelle cet élément a été conçu[68]. Ce genre d'analyse est particulièrement évident face à un artefact inconnu comme le mécanisme d'Anticythère (voir p. 106). Dans le cas qui nous occupe, l'analyse du *design* de la machine révèle que la petite vis a été conçue pour établir une jonction entre deux de ses parties, car le fonctionnement de l'ensemble dépend de cette jonction. Et peu importe que le concepteur de la machine n'en ait pas été conscient. La vis a une *raison d'être* là où elle est, indépendamment des causes (circonstancielles) de sa présence. Et sa fonction est sa raison d'être, comme dirait Ayala (1977). L'analyse fonctionnelle entendue comme une ingénierie inverse permet donc de comprendre pourquoi la vis *doit être* là où elle est, mais ne fournit pas une explication causale de sa présence. Le même type d'analyse et les mêmes conclusions s'appliquent aux valves cæcales des lézards de Pod Mrčaru.

Pour comprendre le fonctionnement d'un artefact, il faut associer son organisation et son comportement avec les intentions présumées de son créateur et les circonstances de l'époque. Par exemple, s'interroger sur la structure du mécanisme d'Anticythère revient à se demander pourquoi (dans quelle but) les Grecs l'ont fabriqué. De façon générale, on peut déduire les intentions du créateur à partir de l'observation de l'objet lui-même. Comme dit Daniel Dennett (1990, p. 182), il n'est pas nécessaire de consulter la biographie d'Alexandre Graham Bell pour savoir ce qu'est un téléphone ni pourquoi (dans quel but) il a été créé. Mais les fonctions (et les intentions) originales ne sont pas toujours aussi évidentes. C'est le cas lorsque le système est cassé ou incomplet, comme le mécanisme d'An-

68 Dans le domaine de la biologie, les capacités pertinentes sont liées aux pressions sélectives auxquelles les organismes sont soumis, et l'analyse procède en considérant ces pressions avec une précision croissante, des plus générales aux plus concrètes : « *One starts from the most general evolutionary pressures, stemming from the competition to reproduce and concomitant needs to survive to sexual maturity, to produce gametes, to identify and attract mates, and so forth. In the context of general features of the organisms in question and of the environments they inhabit, we can specify selection pressures more narrowly, recognizing needs o process certain types of food, to evade certain kinds of predators, to produce particular types of signals, and so forth.* » (Kitcher, 1993, p. 262)

ticythère, de sorte qu'il ne produit pas les effets pour lesquels il a été créé. C'est aussi le cas lorsque les effets qu'il produit n'ont pour nous aucune signification, aucune valeur ; on sait ce que l'appareil fait ou ce que l'on peut faire avec, mais on ne voit pas en quoi cela justifie son existence. Par exemple, pour comprendre la fonction des artefacts d'une culture très différente de la nôtre, il faut les replacer dans leur contexte, car c'est relativement aux usages, aux besoins et aux croyances de la culture à laquelle ils appartiennent que ces artefacts prennent un sens et une valeur qui justifie leur existence. Lorsque la culture d'origine de l'objet nous est inconnue, on peut cependant essayer d'imaginer à quels besoins (ou désirs) il pouvait bien répondre : à quoi pouvaient bien servir les alignements mégalithiques de Stonehenge en Angleterre ou les figures de Nazca au Pérou ? Si un objet ne répond à aucun besoin, s'il n'a aucune valeur, ne serait-ce qu'esthétique, alors il n'y a apparemment aucune raison de le créer. Pour comprendre l'existence et la fonction d'un trait organique comme pour comprendre celle d'un artefact, on peut s'interroger sur les besoins auxquels il répond[69], c'est-à-dire sur ce qui fait sa valeur pour l'organisme.

4.2. Optimalité et bricolage

Un même objet peut répondre à des besoins très différents. Un téléphone portable peut faire office de presse-papiers, un canon peut servir à tuer des mouches et un marteau-piqueur à planter des clous, mais nous ne dirions pas que c'est la fonction pour laquelle ils ont été conçus. Pourquoi ? Parce qu'ils sont peu efficaces, peu efficients ou trop complexes pour l'usage qu'on leur donne. Comme norme générale, il est raisonnable de penser que la fonction d'un artefact (ce pour quoi il a été conçu) est ce pour quoi il est le mieux adapté, car c'est ce que ferait un ingénieur rationnel. Et bien que les ingénieurs ne soient pas parfaits et commettent parfois des erreurs stupides, la présupposition d'optimalité semble néanmoins être un bon critère d'attribution fonctionnelle par défaut (Dennett, 1995, p. 212-213).

Cela étant, comme disait François Jacob (1977), la sélection naturelle — ou, plus généralement, l'évolution biologique — n'opère pas comme un ingénieur mais comme un bricoleur qui fait ce qu'il peut avec les moyens du bord, recyclant de vieux objets pour de nouveaux usages et sans savoir à l'avance ce qu'il va faire ni par quelles étapes il va passer. L'une des conséquences de cette façon de faire, que ce soit en biologie ou

69 La perspective centrée sur l'environnement qu'adopte Kitcher consiste à dire que ces besoins sont les pressions sélectives qui s'exercent sur l'organisme. Cette perspective, dit-il, a des affinités évidentes avec l'idée que les organismes font face à des « problèmes » sélectifs posés par l'environnement (Kitcher, 1993, p. 262). L'idée a été critiqué par Lewontin, et Kitcher répond à ses critiques.

dans le domaine des artefacts, est la présence d'items ou de traits non-fonctionnels (vestigiaux) qui n'ont, par conséquent, aucune raison d'être si ce n'est qu'ils sont liés à d'autres traits ou items qui, eux, sont fonctionnels. L'autre conséquence évidente est la sub-optimalité des résultats, car on peut penser que le bricoleur ne cherche pas l'optimalité mais seulement la satisfaction de ses fins sous certaines conditions.

Cela remet-il en question l'optimalité comme critère d'attribution fonctionnelle ? Non, car ce critère n'implique pas que l'objet en question soit (parmi toutes les alternatives possibles) le mieux adapté pour réaliser une fonction donnée, mais que cette fonction (parmi d'autres possibles) est celle pour laquelle l'objet est le mieux adapté. Ainsi, une ampoule électrique à filament est sub-optimale en tant que dispositif d'éclairage, car elle chauffe plus qu'elle ne brille et a un rendement énergétique très inférieur à celui d'un tube fluorescent ou une lampe à LED, mais elle est restée pendant longtemps le meilleur moyen d'illumination disponible et c'est sans doute là le meilleur usage que l'on pouvait en faire.

De plus, un bricoleur est un agent rationnel. Il est donc raisonnable de penser que s'il a le choix entre plusieurs options pour atteindre ses fins, il choisira celle qui de son point de vue est la meilleure. Or, pour lui, cela peut être la plus rapide, la plus simple ou la plus économique. Si l'on tient compte du coût matériel, temporel, intellectuel, énergétique, économique, etc., alors la solution optimale d'un problème donné n'est pas forcément celle de l'ingénieur qui doit la concevoir et la fabriquer *ex-novo*, mais plutôt celle du bricoleur qui détourne et recycle à moindre coût des éléments pré-existants. Par exemple, un simple rouleau de scotch peut être employé comme source de rayons X grâce au phénomène de la triboluminescence (Camara, Escobar, Hird, & Putterman, 2008). Avec ce même rouleau de scotch et des mines de crayon, les prix Nobel Andre Geim et Konstantin Novoselov ont réussit pour la première fois à isoler des feuillets de graphène pour en étudier les propriétés physiques. Et avec des mines de crayon, de l'eau, du détergent et un mixeur de cuisine, il est possible de fabriquer de grandes quantités de ce matériau (Paton et al., 2014). C'est facile, rapide et économique. Ce n'est peut-être pas la meilleure solution possible, idéalement, mais parmi toutes celles que l'on connaît jusqu'à maintenant, c'est sans doute l'une des meilleures.

Par ailleurs, on peut parfois trouver à un objet un meilleur usage que ce pour quoi il a été créé. Soit parce que le créateur était mal inspiré, soit parce que l'évolution des pratiques et des besoins a découvert un usage possible de cet objet qui n'existait pas au moment de sa conception, ou encore parce que le progrès technique a créé d'autres objets beaucoup plus adaptés ou performants pour le même usage. Si l'attribution fonctionnelle repose exclusivement sur un critère d'optimalité, alors comme dit Dennett (1990), l'inventeur d'un artefact n'est pas l'arbitre final de sa

fonction et doit plutôt être considéré comme un utilisateur parmi d'autres, ses intentions n'ayant à la limite qu'une valeur historique[70].

Nous ne prétendons pas ici contribuer au débat sur les fonctions des artefacts, mais seulement montrer que le bricolage, qu'il soit biologique ou technologique, n'est pas une objection à l'optimalité comme critère d'attribution fonctionnelle par défaut. Que nous soyons face à un artefact dont nous ne connaissons pas l'auteur ni le mode d'emploi, ou face à un trait biologique dont nous ignorons l'histoire sélective, nous sommes néanmoins capables, dans la plupart des cas, d'en « deviner » correctement la fonction, indépendamment du fait que la chose en question ait été produite *ex novo* ou bricolée, cooptée, recyclée à partir de matériaux avec une fonction différente. Nous reviendrons au CHAP. VIII, SECT. 4, sur la question de l'optimalité dans le cadre d'une discussion autour de l'adaptationnisme.

4.3. Optimalité, intentionnalité et finalité

Le critère d'optimalité est-il suffisant pour justifier l'attribution d'une fonction ? Non, car il faut tenir compte de deux autres éléments : la finalité et le contexte (l'environnement). Dans la suite de ce travail, nous allons défendre en effet que les fonctions sont relatives à une fin dans un contexte donné et que c'est dans le cadre de la relation entre ces trois éléments que doit être compris le critère d'optimalité. En d'autres termes, les fonctions sont des moyens efficients d'atteindre une certaine fin dans un contexte donné. Cela signifie que pour déterminer la fonction d'un trait biologique ou d'un item technique, il faut d'abord déterminer la finalité à laquelle il contribue, et cette dernière est généralement liée au contexte.

Par exemple, l'une des fonctions des vêtements en fourrure animale est l'isolation thermique, car ils peuvent contribuer de manière efficiente à conserver la chaleur corporelle, laquelle est nécessaire pour la survie, mais cette fin est liée au contexte climatique des gens qui les portent, c'est-à-dire à un climat froid. Si un anthropologue visitant pour la première fois une tribu sub-tropicale isolée observait des individus porter des manteaux en peau de lion, il leur attribuerait probablement une autre fonction, car le froid n'est pas un problème pour eux. La finalité, dans ce contexte pourrait être de communiquer un certain statut social. Le même type de raisonnement est applicable aux traits biologiques, avec une analogie évidente qui est celle de la fourrure des animaux.

---

70 De même, en littérature, l'auteur d'un texte n'est pas l'arbitre final de sa signification, car un lecteur perspicace peut découvrir des choses dont l'auteur n'était pas conscient. Par exemple, on peut faire une lecture psychanalytique ou féministe ou marxiste de *Don Quichotte* que Cervantes ne pouvait évidemment pas faire.

Dans le cas des artefacts, les fins et les fonctions sont généralement associées à l'intentionnalité du créateur, et donc d'une certaine manière à l'origine causale de l'objet. Notre conception des fonctions est différente et s'appuie sur des relations d'optimalité entre les moyens, les fins et le contexte, mais nous ne pouvons manquer de signaler certaines objections et critiques que l'on peut opposer au principe d'optimalité depuis la psychologie et la philosophie de la technique.

La première est que, selon certaines études (Defeyter & German, 2003; German & Johnson, 2002; Kelemen, 1999b), les personnes tendent à classer les artefacts dans une catégorie fonctionnelle d'après les intentions originales de leur créateur plutôt que selon des critères d'usage et d'optimalité[71]. Dans l'exemple de la petite vis de Kitcher, on voit clairement que l'objet est identifié comme une vis (fonction originale) bien que sa fonction d'usage *dans le cadre de la machine* soit d'établir une jonction électrique.

La seconde raison est que l'attribution de fonctions à des artefacts est sensible aux conventions sociales (voir Vaesen & van Amerongen, 2008). Quand un objet est utilisé différemment de ce pour quoi il a été conçu, plus le nouvel usage est culturellement établi et plus nous aurons tendance à le reconnaître comme sa fonction propre, surtout si l'histoire de l'objet et les intentions originales de son créateur nous sont inconnues. Ainsi, même si nous apprenions que l'épluche-patate a initialement été conçu comme un outil pour récolter le caoutchouc, nous continuerions à lui attribuer la fonction de peler les légumes (Perlman, 2009, p. 61) tout en reconnaissant que *dans le cadre des plantations de caoutchouc* le même objet a une fonction différente.

La troisième raison est que, même si un objet s'avère optimal pour un usage donné, on ne lui attribuera pas forcément la fonction correspondante si ce n'est pas un artefact. La planète Jupiter peut être utilisée comme un tremplin gravitationnel pour envoyer des sondes spatiales vers les confins du système solaire en minimisant l'énergie embarquée, et c'est peut-être la meilleure chose qu'on puisse faire avec, mais ni les planètes ni leurs champs gravitationnels n'ont de fonctions. De même, on n'attribuera pas de fonction à un objet ni à ses parties s'ils sont des produits du hasard. Si deux explorateurs découvrent en montagne un amas rocheux qui ressemble à un cadran solaire à tel point qu'ils peuvent l'utiliser comme tel, ils ne diront pas pour autant que le rocher vertical est un gnomon dont la fonction est de projeter une ombre indiquant la position du soleil. Cela étant, un objet naturel peut être utilisé comme si c'était un artefact et être considéré comme tel indépendamment de son origine causale (Bloom, 1996). En effet, si nos deux explorateurs découvrent que

71 Plusieurs autres études (voir Vaesen & van Amerongen, 2008, p. 13) ont montré que les caractéristiques physiques de l'objet peuvent prévaloir sur les intentions originales, mais sans que ces dernières soient pour autant écartées.

l'amas rocheux porte des traces de peinture compatibles avec son utilisation comme cadran solaire par les populations locales et qu'il a été pendant des siècles le centre d'une activité sociale orientée dans le même sens, alors ils pourront pointer du doigt le rocher vertical et le décrire comme un gnomon. L'amas rocheux est évidemment le même, mais tandis que dans le premier cas les randonneurs interprètent (correctement) sa capacité à indiquer l'heure comme un produit du hasard, ils peuvent dans le deuxième cas le considérer comme un artefact *dans le cadre des pratiques culturelles des populations locales* et lui attribuer la fonction correspondante comme s'il avait été créé intentionnellement pour cela. Le critère d'optimalité s'applique également dans les deux cas, car l'objet considéré est le même ; pourtant, l'attribution de fonctions à la manière de l'ingénierie inverse n'est possible que dans le second[72].

Dans le domaine de l'archéologie, la catégorisation des artefacts et l'attribution de fonctions est souvent relative aux intentions — présumées — de leurs créateurs. Vaesen et van Amerongen (2008, p. 16) citent l'exemple d'un vase babylonien en argile contenant des cylindres de cuivre et de fer emboîtés et isolés l'un de l'autre par du bitume (Fig. 7). En 1938, l'objet a été interprété par Wilhelm König, directeur du Musée National d'Irak, comme étant une cellule galvanique. La confirmation de cette hypothèse signifierait que l'électricité a été découverte en Mésopotamie plus de 1800 ans avant Galvani et Volta. En 1993, l'hypothèse de König a été testée avec succès : il a été démontré que le vase peut produire une faible tension électrique lorsqu'il est rempli d'une solution acide (jus de citron, vinaigre). C'est-à-dire qu'en se limitant à un critère d'optimalité, l'interprétation de cet objet en tant que batterie primitive serait correcte et on pourrait alors attribuer à ses éléments les fonctions correspondantes. Les partisans du paranormal et des visites d'extraterrestres applaudiraient chaudement. Cela étant, ce que les expériences montrent, c'est seulement une possibilité physique, pas un usage réel. De fait, aucun usage de l'électricité n'a pu être vérifié à cette époque. Les

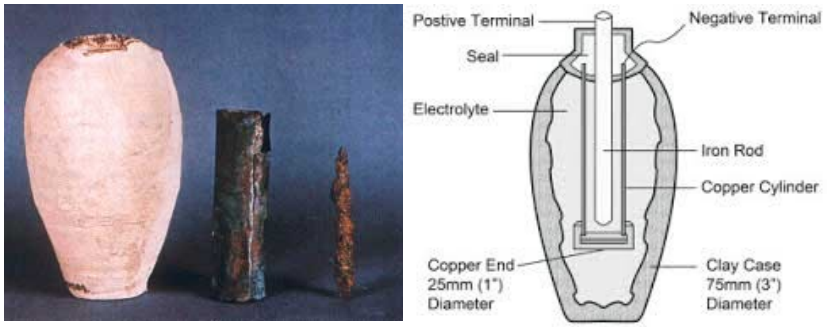

Figure 7: La « batterie babylonienne ».

72 Par contre, on peut toujours réaliser une analyse fonctionnelle à la Cummins.

archéologues sceptiques penchent plutôt vers une interprétation plus raisonnable, à savoir que le vase — similaire à ceux trouvés à Séleucie du Tigre — servait à conserver des rouleaux sacrés. Le problème est que pour entreposer un parchemin, n'importe quel récipient ferait l'affaire, or celui-ci est très — voire trop — complexe. Pour rendre compte de la présence des différents éléments, il faut alors faire référence aux désirs et aux croyances des utilisateurs de l'époque pour qui le cuivre et le fer avaient peut-être une signification symbolique. Appliqué à un artefact, le critère d'optimalité est donc un excellent outil heuristique lorsqu'on ne connaît pas les intentions de son créateur ou de ses utilisateurs, mais il ne permet pas de se passer de l'intentionnalité.

La relation entre l'intentionnalité et la fonctionnalité n'est pas la même selon que l'on considère un artefact entier ou seulement l'une de ses parties. L'argument de Kitcher à ce propos, notamment à travers l'exemple de la petite vis, est que la fonction des parties d'un artefact est relative à leur contribution au fonctionnement de la machine dans son ensemble, tandis que la fonction de cette dernière est relative aux intentions du créateur. Autrement dit, une fois que la fonction de la machine est intentionnellement fixée, la fonction de ses parties n'a pas besoin d'être connue du créateur et elle peut-être déterminée par rétro-ingénierie. Par exemple, si les mésopotamiens avaient conçu — ou seulement utilisé — la « batterie de Bagdad » pour faire de la galvanoplastie, on pourrait analyser le rôle causal de ses parties et leur attribuer des fonctions correspondantes à celles d'une vraie batterie, indépendamment des connaissances en chimie et en électricité de ceux qui l'ont fabriquée.

On peut remplacer la référence à l'intentionnalité par une référence à la finalité. L'un des avantages de cette substitution est qu'elle nous permet de mieux distinguer la fonction d'un artefact (entendue comme ce pour quoi il a été conçu ou est utilisé : sa finalité) de la fonction de ses parties (entendue comme leur contribution causale à la réalisation de cette fin).

Un autre avantage de cette substitution est qu'elle nous permet de mieux comprendre le recours à l'analyse fonctionnelle dans le cadre d'une approche comme celle de Kitcher. Au lieu de prendre comme point de départ de l'analyse une capacité quelconque d'un système, à la manière de Cummins, on peut partir de la finalité générale de ce système, analyser la contribution causale de ses éléments relativement à la réalisation de cette finalité et leur attribuer la fonction correspondante suivant un critère d'optimalité.

Un troisième avantage de cette substitution est qu'elle nous permet de généraliser au domaine de la biologie les conclusions précédentes relatives aux artefacts. Si l'on entend que la finalité générale d'un organisme biologique est la survie et la reproduction, alors on pourra interpréter les fonctions de ses parties en termes de contribution à la réalisation de cette fin. Cette contribution passe en particulier, suivant l'approche de Kitcher,

par la réponse aux pressions sélectives de l'environnement. Ainsi, les valves cæcales des lézards de Pod Mrčaru ont une fonction dans la mesure où elles contribuent à leur survie en facilitant la digestion de la cellulose dans un environnement pauvre en insectes et riche en plantes.

En attribuant une finalité aux organismes biologiques et en interprétant l'analyse fonctionnelle à la manière d'une rétro-ingénierie, on peut attribuer des fonctions à leurs parties indépendamment de leur origine causale, c'est-à-dire indépendamment du fait qu'ils soient le produit d'une création intelligente ou du mécanisme de sélection naturelle.

Conclusions de la première partie

Les différentes tentatives de naturalisation des fonctions biologiques sont à la fois séduisantes et insatisfaisantes. Séduisantes parce qu'elles capturent des aspects intuitivement importants de la notion de fonction. Insatisfaisantes parce que chacune d'elles laisse de côté d'autres aspects non moins importants que retiennent les autres. Insatisfaisantes aussi parce qu'elles sont confrontées à des contre-exemples et des objections tenaces qui les rendent difficilement acceptables.

Les approches étiologique et systémique ne sont pas concurrentes ni radicalement distinctes, mais complémentaires. Elles analysent la même notion depuis des perspectives différentes qui correspondent à des besoins et des usages différents en biologie évolutive et en biologie fonctionnelle. Nonobstant, il semble possible de rassembler les éléments essentiels de ces deux approches pour proposer une conception unifiée qui s'appuie sur une forme d'ingénierie inverse applicable aussi bien aux fonctions biologiques que techniques.

Pour arriver à une telle conception, il faut laisser de côté certains présupposés ontologiques et épistémologiques courants. En premier lieu, il faut accorder une place aux valeurs dans l'approche étiologique et aux normes dans l'approche systémique. Cela implique qu'il faut soit admettre que les valeurs et les normes existent dans la nature, soit abandonner l'idée que les fonctions sont des propriétés indépendantes de l'observateur. En second lieu, il faut aussi reconnaître que les explications fonctionnelles ne sont pas causales, mais téléologiques, ce qui veut dire qu'elles s'inscrivent dans une relation de type moyens–fin.

L'une des questions soulevées par cette analyse du débat porte sur le lien entre les fonctions et la nature des objets auxquels on les attribue. Autrement dit, pourquoi les parties des êtres vivants peuvent elles être fonctionnelles, mais pas celles des systèmes naturels non biologiques ? C'est l'une des questions que nous aborderons dans la seconde partie.

Deuxième partie :
LES FONCTIONS AUX FRONTIÈRES DU VIVANT

[M]ême les plus ouvertement réductionnistes des biologistes contemporains trouvent un sens à la question des frontières entre le vivant et l'inerte.

Anne Fagot-Largeault, « Le vivant » (1995, p. 245)

Il n'est pas évident que les êtres vivants soient une catégorie naturelle, et non une catégorie humaine, le fruit d'une décision commune de classer les êtres vivants dans une catégorie à part.

Michel Morange, « La principale difficulté… » (2007, p. 65)

My attitude, which might be labelled empirical nihilism, is that the statement that a system is or is not alive is a statement about the speaker's attitude of mind rather than about the system, and that no question is scientifically relevant unless the questioner has an experiment in mind by which the answer could be approached

Norman Pirie, « The origins of life » (1957)

INTRODUCTION DE LA DEUXIÈME PARTIE

L'objectif principal de cette deuxième partie est d'éclairer le débat sur la naturalisation des fonctions à la lumière des recherches en biologie synthétique et sur les origines de la vie. En effet, ces recherches visent notamment à décrire et à produire des cellules minimales, c'est-à-dire les formes de vie les plus simples possibles. Or, si les objets biologiques ont des fonctions tandis que les autres systèmes naturels n'en ont pas, alors l'étude des frontières et des conditions de possibilité du vivant devrait nous renseigner aussi sur les frontières et les conditions de possibilité des fonctions biologiques. Ces recherches pourraient par exemple nous aider à comprendre comment les fonctions d'un système apparaissent lorsqu'il passe du statut de processus chimique prébiotique à celui d'organisme vivant à part entière. Il ne s'agit pas ici de savoir ce que les spécialistes *pensent* à propos des fonctions, mais d'essayer de comprendre ce qu'elles *sont*.

Nous commencerons au CHAP. IV par examiner les deux grandes approches suivies en biologie synthétique. La première, réductive (*top-down*), consiste à simplifier le génome d'organismes actuels jusqu'à atteindre l'ensemble minimal de gènes nécessaires pour que la cellule accomplisse ses fonctions de base. La seconde, constructive (*bottom-up*), a pour objectif de comprendre le passage de l'inerte au vivant à partir de processus physico-chimiques simples comme ceux ayant dû se produire sur Terre il y a près de quatre milliards d'années.

Nous verrons ensuite que les tentatives de naturalisation du vivant rencontrent des problèmes d'identification et de caractérisation de leur objet analogues à ceux rencontrés par la naturalisation des fonctions, car les stratégies adoptées sont similaires.

Nous nous interrogerons au CHAP. V sur la réalité et sur l'objectivité scientifique du vivant et de ses frontières, car si la distinction entre le vivant et l'inerte dépendait en partie de l'observateur, il pourrait en être de même des fonctions biologiques. Ce qui pour nous est en question, ce n'est pas le fait que les phénomènes qui relèvent de ce que nous appelons la vie soient explicables ou non par les sciences de la nature, car nous supposerons qu'ils le sont. Ce n'est pas non plus l'existence de cas limites, intermédiaires ou indéfinis entre la vie et la non-vie, car l'absence d'une

distinction bien tranchée ou l'existence d'effets de bord entre deux types de phénomènes naturels ne veut pas dire qu'ils ne soient pas distincts.

Ce qui pour nous est en question, premièrement, c'est le fait qu'il y ait ou pas une unité sous-jacente à la vie prise dans son universalité. C'est-à-dire le fait que l'ensemble des objets que nous reconnaissons comme vivants, dans toute la diversité de leurs manifestations possibles, forment ou pas une collection non-arbitraire d'individus.

Ce qui est en question, deuxièmement, c'est la nature des principes, lois ou propriétés sur lesquelles repose cette unité présumée du vivant. Ont-elles une efficacité causale ? Sont-elles ontologiquement objectives ou subjectives ? Intrinsèques ou relatives ? S'agit-il de qualités premières ou secondes ?

Ce qui est en question, troisièmement, c'est le fait que les êtres vivants, en tant que tels, ne constituent pas seulement un type d'objets physicochimiques parmi d'autres, une simple branche dans la taxonomie universelle de la physique, mais quelque chose d'autre, un type d'objets à nul autre pareil.

Chapitre IV

Fabriquer, mesurer, classer

1. La biologie synthétique et les cellules minimales

On assiste depuis une quinzaine d'années à l'émergence d'un nouveau champ de recherches visant à produire et à contrôler des objets — les plus simples possibles — pouvant être considérés comme vivants[73]. Selon certains auteurs (Luisi, 1998; Luisi, Chiarabelli, & Stano, 2006; Szathmary, 2005; Szostak, Bartel, & Luisi, 2001) ces recherches nous permettront de mieux comprendre les origines de la vie sur Terre et de la reproduire en laboratoire. Elles nous permettront non seulement d'identifier et de situer la frontière qui sépare le vivant du non-vivant, mais aussi et surtout de saisir ses conditions de possibilité et comprendre ses mécanismes les plus élémentaires. « *Créer de la vie* ex novo, disent Alonso Ricardo et Jack Szostak (2009), *nous aidera sans doute à comprendre comment la vie peut commencer, quelle est la probabilité qu'elle existe sur d'autres mondes et, en définitive, ce qu'est la vie* ».

Dans la mesure où elles explorent les conditions de possibilité du vivant, ces recherches devraient également nous renseigner sur les conditions de possibilité des fonctions biologiques (Bedau, 2007; Lee, Severin, & Ghadiri, 1997; Pohorille, 2011; Rasmussen, Chen, Nilsson, & Abe, 2003; Szostak, 2008). Elles pourraient, par exemple, nous permettre de comprendre comment les fonctions d'un système apparaissent lorsque

73 La réflexion sur la définition de la vie dans le cadre de la biologie synthétique a été initialement menée en collaboration avec Manouk Abkarian (Laboratoire Charles Coulomb UMR 5221 CNRS-UM2, Montpellier). Certains résultats de cette réflexion ont été présentés à l'occasion d'un congrès (Molina Pérez & Abkarian, 2007). Certaines conclusions ultérieures ont été publiées par ailleurs (Molina Pérez, 2009).

celui-ci passe du statut de processus chimique prébiotique à celui d'organisme vivant à part entière.

Ce qui intéresse certains chercheurs et les institutions qui les financent, ce n'est pas tant le statut vital des entités créées que les fonctions que nous serons capables de leur faire remplir. Car la biologie synthétique promet d'un côté de réduire les organismes vivants à leur plus simple expression en identifiant les constituants d'une cellule minimale, et, d'un autre côté, de créer des formes de vie dotées de fonctions nouvelles par le biais de la reprogrammation génétique, de l'assemblage de « Légos » moléculaires (*BioBricks*™) ou de protéines inédites. C'est-à-dire que la biologie synthétique est aussi et surtout une discipline biotechnologique (Benner & Sismour, 2005; Jaffe, 2005; Serrano, 2007) intéressée — dans tous les sens du terme — par les fonctions biologiques.

Au-delà de leur intérêt théorique, ces recherches présentent un intérêt technologique majeur, qui est la maîtrise d'unités de production biochimiques à haut rendement, versatiles, quasi-autonomes, se prêtant aussi bien à des applications extrêmement locales qu'à une industrialisation massive. Craig Venter n'est sans doute que le premier d'une longue liste à tenter de breveter un être vivant artificiel, *Mycoplasma laboratorium*.

Pour obtenir une cellule minimale on peut procéder de deux manières : par réduction (*top-down*) et par construction (*bottom-up*). L'approche réductive consiste à simplifier le génome d'organismes actuels jusqu'à atteindre l'ensemble minimal de gènes nécessaires pour que la cellule accomplisse ses fonctions de base et au-delà duquel l'organisme résultant ne peut plus être considéré comme vivant. C'est la méthode suivie par Craig Venter et son équipe à partir de la bactérie *Mycoplasma genitalium* dont ils ont identifié comme essentiels 382 des 482 gènes codant pour des protéines (Glass et al., 2006). Certaines estimations situent la limite inférieure entre 50 et 300 gènes (Luisi, Ferri, & Stano, 2006; Szathmary, 2005).

La compréhension et le contrôle de cellules vivantes aussi simples que possible est extrêmement utile pour développer et systématiser l'ingénierie du vivant. L'un des objectifs de Craig Venter est en effet la synthèse de bactéries artificielles dont le code génétique serait programmable à volonté (Gibson et al., 2008, 2010). D'autres équipes s'affrontent lors de compétitions organisées par le Massachusetts Institute of Technoloy (MIT) pour produire toutes sortes de dispositifs biologiques artificiels à partir d'éléments fonctionnels élémentaires de la machinerie du vivant (Check, 2005). Un registre de parties biologiques standardisées appelées *BioBricks*™[74] a d'ailleurs été mis en place par le MIT pour faciliter la

74 « We define a biological part to be a natural nucleic acid sequence that encodes a definable biological function, and a standard biological part to be a biological part that has been refined in order to conform to one or more

construction de systèmes vivants à partir de pièces interchangeables (Endy, 2005; Gibbs, 2004; Sprinzak & Elowitz, 2005). Parallèlement à la technique « Frankenstein », qui consiste à assembler des parties d'organismes aux fonctions connues pour donner vie à une créature non naturelle, d'autres explorent les possibilités non réalisées de la nature pour y découvrir des fonctions inédites ou pour reproduire autrement — plus efficacement (?) — des fonctions biologiques connues (Benner & Sismour, 2005). Il s'agit notamment de synthétiser de façon aléatoire des protéines totalement nouvelles (*never born proteins*) pour ensuite sélectionner et isoler celles qui ont un potentiel intéressant (Luisi, 2006, 2007; Luisi, Chiarabelli, et al., 2006). Toutes ces techniques sont réductives dans la mesure où elles partent de cellules vivantes déjà constituées qu'elles analysent, décomposent, recomposent, transforment, imitent et réinventent ; mais personne n'a pour l'instant réussi à fabriquer une cellule entière à partir de ses constituants élémentaires[75].

Ces techniques reposent toutes sur une sorte d'ingénierie génétique inverse et assument de ce fait certaines idées familières aussi bien aux biologistes systémiques qu'aux ingénieurs mais pas forcément intelligibles du point de vue d'un physicien. Parmi elles, la métaphore cartésienne d'une machinerie cellulaire et l'idée de fonction associée à ses parties[76]. En effet, l'idée fondamentale derrière l'approche réductive est que « *n'importe quel système biologique peut-être vu comme un assemblage d'éléments fonctionnels individuels — a l'image de ceux qu'on trouve dans les appareils manufacturés* » (de Lorenzo & Danchin, 2008, p. 822).

L'approche constructive vise à fabriquer artificiellement les systèmes chimiques les plus simples possibles pouvant être considérés comme vivants. Elle a notamment pour objectif de comprendre l'apparition de la vie à partir de processus physico-chimiques similaires à ceux ayant dû se produire sur Terre il y a quatre milliards d'années (Luisi, 1998; Luisi,

---

defined technical standards. […] Parts that conform to the BioBrick assembly standard are BioBrick standard biological parts. » (Shetty, Endy, & Knight, 2008)

75 On sait depuis les années 1950 que dans certaines conditions le virus de la mosaïque du tabac se reconstitue spontanément *in vitro* à partir de ses constituants : le génome isolé et les protéines capsulaires. Luisi (2006, p. 5) rapporte que des travaux postérieurs ont réussi à reconstituer des noyaux cellulaires et des cellules entières — amibes et algues vertes — à partir du noyau, du cytoplasme et de la membrane cellulaire. Toutefois, ajoute cet auteur, les réassemblages en question ne sont pas des processus spontanés et doivent être « guidés » par micro-chirurgie. De plus, les éléments isolés employés pour la reconstitution de cellules vivantes ne sont pas des constituants moléculaires mais des systèmes qui sont eux-mêmes déjà hautement organisés.

76 C'est-à-dire l'idée non problématisée selon laquelle les parties d'un organisme ont des fonctions.

Chiarabelli, et al., 2006; Szathmary, 2005; Szostak et al., 2001). Ces recherches sont un prolongement des fameuses expériences de Stanley Miller (1953) sur la synthèse d'acides aminés dans la « soupe primitive ».

La notion de vie, depuis cette perspective, n'est pas donnée. On l'explique souvent en termes de propriétés émergentes (Luisi, 2002; Szostak et al., 2001). Les publications dans le cadre de la méthode constructive commencent souvent par une définition de la vie ou par une liste de critères servant de justification au travail expérimental de leurs auteurs. La méthode consiste alors à créer les objets les plus simples possibles qui soient conformes à la définition donnée. Il s'agit généralement de vésicules auto-réplicantes dans lesquelles les chercheurs introduisent des polymères servant de support d'information ainsi qu'un système métabolique primitif (Orgel, 1992; Szathmary, 2005). La possession d'un code génétique à base d'ADN ou d'ARN n'est pas jugée nécessaire par tous les auteurs, car les formes de vie les plus simples et les plus primitives ont pu apparaître avant la formation de ces molécules complexes.

On retrouve dans les méthodes exposées ci-dessus deux conceptions complémentaires de la vie. Une conception systémique centrée sur la complexité intrinsèque des organismes et une conception darwinienne centrée sur la réplication et l'évolution[77]. Les deux apparaissent conjointement dans les critères du vivant mentionnés par la plupart des chercheurs en biologie synthétique, en exobiologie et dans le domaine des origines de la vie. Certains d'entre eux (Luisi, Ferri, et al., 2006; Rasmussen et al., 2004) considèrent en effet qu'un système moléculaire est vivant s'il est capable de s'auto-régénérer continuellement, de s'auto-répliquer et d'évoluer. Cette conception, qui rejoint celle de la NASA (Joyce, 1994), a été renforcée par l'idée du monde à ARN permettant de résoudre le paradoxe de l'œuf et de la poule entre l'origine des protéines (et du métabolisme) et celle des acides nucléiques (l'information génétique) (Chyba & McDonald, 1995; Orgel, 2004). D'autres auteurs (Gánti, 2003; Morowitz, 1993; Shapiro, 2007) ajoutent comme condition nécessaire la présence d'une membrane, quitte à laisser entre parenthèses la condition métabolique (Szostak et al., 2001). La double conception systémique et historique est par ailleurs conforme à la théorie de Freeman Dyson (1985) de la double origine de la vie sous la forme d'un monde de protéines (capable d'activité métabolique mais pas de réplication) et d'un

77 « [Il existe] une division assez stricte entre ceux qui pensent que la vie est un système interactif émergent doté de propriétés dynamiques qui existe dans un état dont le comportement est proche d'un comportement chaotique et ceux qui ne peuvent adhérer à la définition d'un système vivant sans une composante génétique dont les propriétés reflètent le rôle de la sélection naturelle de type darwinien et plus généralement le rôle des processus évolutionnistes » (Lazcano, 2007, p. 52). Pour un bref historique des deux conceptions, voir Lazcano (2008) qui reprend et complète l'article précédent.

monde postérieur à ARN (capable de réplication mais pas de métabolisme). Elle rejoint également l'idée de la double nature défendue par John Maynard Smith et Eors Szathmary (1999).

2. Des cellules minimales à la vie minimale

Les recherches en biologie synthétique et sur les cellules minimales pourraient apparaître comme un modèle d'unité des sciences dans la mesure où elles font converger plusieurs champs disciplinaires distincts vers un objet commun se situant à l'interface de la biologie et de la physique, à l'endroit même où, sur une échelle de complexité, la matière organisée devient organisme vivant (Fig. 8).

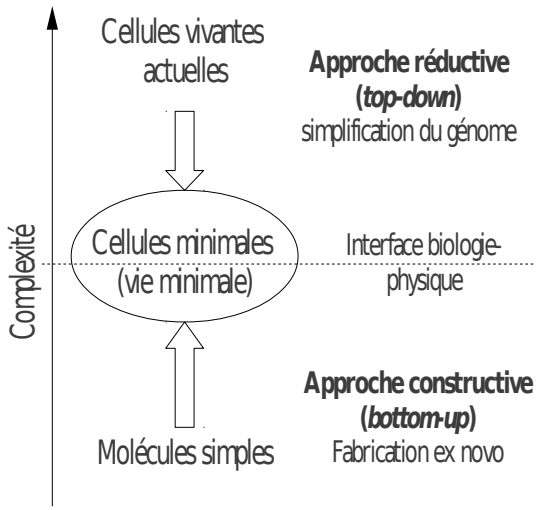

Figure 8: Cellules minimales : un objet transdisciplinaire ?

Pourtant, à y regarder de plus près, les deux approches ne semblent pas devoir aboutir au même point. En premier lieu parce que, quand bien même le nombre minimal de gènes nécessaires à une cellule vivante serait inférieur à la centaine, cela représente encore une complexité considérable dont on peut difficilement croire qu'elle fut celle des premières formes de vie sur Terre (Szathmary, 2005), d'autant plus que la méthode réductive ne nous permet pas de comprendre comment ces organismes auraient pu surgir à partir d'entités non vivantes génétiquement plus simples. On peut raisonnablement penser que cette approche ne nous ramène pas aux origines de la vie elle-même mais à celles de la vie *telle que nous la connaissons*. L'approche constructive présente quant à elle l'avantage de montrer comment se fait la transition du non-vivant vers le vivant à partir de processus physicochimiques simples, mais tout en nous présentant les mécanismes possibles de cette transition, elle ne nous

garantit pas de trouver le bon, c'est-à-dire celui qui a été à l'œuvre il y a presque quatre milliards d'années.

En second lieu, entre les objets produits par chacune des deux approches la différence de complexité est assez considérable. La cellule vivante à ADN la plus simple actuellement envisageable compte approximativement une cinquantaine de gènes (Luisi, Oberholzer, & Lazcano, 2002) tandis qu'une hypothétique cellule vivante à ARN pourrait n'être composée en tout et pour tout que d'une membrane lipidique auto-reproductive (par croissance et division spontanée) et de deux ribozymes, l'une capable de synthétiser la membrane et l'autre capable de se répliquer elle-même et de répliquer la première ribozyme (Monnard, 2011; Oberholzer, Wick, Luisi, & Biebricher, 1995; Szostak et al., 2001).

Les critères qui sous-tendent l'attribution de vie à ces deux types de cellules ne sont pas les mêmes. La première est considérée comme vivante dans la mesure où elle est capable d'homéostasie, d'auto-reproduction et d'évolution. La seconde, capable seulement d'auto-réplication et de mutation, serait conforme à la définition de la NASA (Joyce, 1994; Luisi, 1998). Se pose alors la question de savoir laquelle de ces deux listes de critères correspond effectivement aux origines de la vie, sachant par ailleurs que les cellules à ARN auraient pu évoluer progressivement jusqu'à aboutir aux cellules à ADN telles que nous les connaissons.

Si l'auto-réplication et l'évolution sont suffisantes pour qualifier une cellule de « vivante », alors le même qualificatif devrait être applicable aux populations de molécules de l'hypothèse d'un monde à ARN, car la définition de la NASA n'exige pas leur encapsulation dans une membrane[78]. Et l'ARN étant une molécule très complexe, elle a probablement évolué à partir d'autres polymères comme l'acide nucléique peptidique ou l'acide nucléique à thréose, ou à partir de molécules informationnelles plus simples (Orgel, 1992, 1998, 2004). À ce propos, Luisi (2006, p. 56) va même jusqu'à dire que l'idée — répandue chez les biologistes moléculaires — selon laquelle des molécules d'ARN auto-répliquant seraient apparues spontanément à partir de la soupe prébiotique est comparable à la croyance en un miracle. Cela nous conduit à élargir encore le fossé entre les approches réductive et constructive en repoussant la frontière de la vie minimale vers ce qu'on pourrait appeler une éventuelle « vie moléculaire » constituée par exemple d'écosystèmes moléculaires, c'est-à-dire de réseaux cycliques auto-organisés de molécules auto-réplicantes (Lee et al., 1997; Lincoln & Joyce, 2009).

78 La membrane n'est pas nécessaire si la concentration de molécules est suffisante ou s'il existe une autre forme de compartimentation, par exemple minérale. La vie aurait ainsi pu apparaître dans les interstices de feuillets de mica, selon l'hypothèse de Helen Hansma. Voir aussi note 111, p. 192.

Par ailleurs, il est possible de repousser vers le bas la frontière des cellules minimales en remplaçant l'ADN et l'ARN — ou n'importe quelle autre molécule informationnelle — par des processus physicochimiques beaucoup plus simples. Nous savons actuellement fabriquer des vésicules lipidiques[79] capables non seulement d'auto-reproduction (Bachmann, Walde, Luisi, & Lang, 1990; Luisi, 2006) mais aussi d'homéostasie (Walde, Wick, Fresta, Mangone, & Luisi, 1994; Zepik, Bloechliger, & Luisi, 2001). Le modèle le plus simple est un système constitué d'un conteneur et de deux réactions chimiques concurrentes, la première crée les éléments du conteneur tandis que la seconde les détruit (Fig. 9). Le conteneur est une membrane semi-perméable formée de molécules *S* (surfactants) laissant passer les molécules *A* présentes dans le milieu. À l'intérieur du système, les molécules *A* se transforment en molécules *S* qui s'ajoutent à la membrane. De façon parallèle, ces dernières se dégradent en molécules *P* qui retournent dans le milieu. Lorsque les vitesses de génération et de dégradation de la membrane sont égales, le système atteint l'homéostasie. Lorsque la vitesse de génération est supérieure à celle de dégradation, le système peut croître et s'auto-reproduire[80]. Un tel système peut être considéré comme vivant dans la mesure où il manifeste trois propriétés essentielles partagées par tous les êtres vivants connus : il est auto-contenu, auto-généré et auto-maintenu (Fleischaker, 1990). Il satisfait par conséquent les conditions de la définition autopoïétique de la vie (Luisi, 2003; Luisi & Varela, 1989).

79 Les vésicules et les liposomes constituent de bons modèles pour comprendre les membranes cellulaires, car ils possèdent des propriétés physicochimiques analogues et pourraient, de fait, avoir contribué à l'émergence de la vie (Luisi, Walde, & Oberholzer, 1999).

80 À partir d'une certaine taille, les vésicules lipidiques peuvent devenir instables et se diviser spontanément en vésicules filles dotées des mêmes propriétés que la vésicule mère.

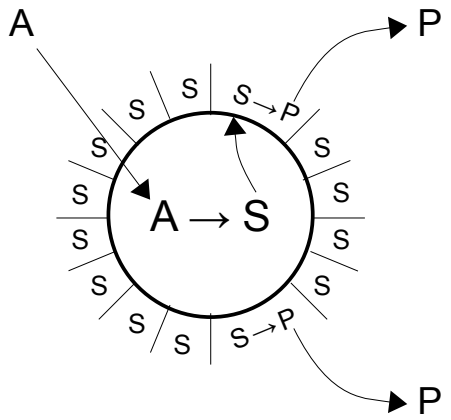

Figure 9: Système autopoïétique minimal. Ce système se caractérise par deux réactions compétitives, l'une construisant les éléments constitutifs de la membrane tandis que l'autre les détruit. Selon la vitesse relative de ces deux réactions, le système peut se maintenir en homéostasie, grandir ou mourir. (Source : Luisi, 2003).

Dans sa formulation initiale (Maturana & Varela, 1973; Varela, Maturana, & Uribe, 1974), la définition autopoïétique ne mentionne ni l'évolution ni la reproduction parmi les conditions nécessaires du vivant. De fait, contrairement aux cellules minimales à ADN, aux cellules réplicantes à ARN et aux écosystèmes moléculaires, les micelles autopoïétiques minimales ne contiennent aucun support d'information permettant une évolution darwinienne. Si la reproduction et l'évolution restent possibles[81], elles se font sans hérédité, par simple copie/duplication du système.

Nous nous retrouvons par conséquent avec plusieurs origines potentielles et plusieurs formes de vie minimale possibles sans lien direct entre elles, correspondant à des définitions différentes et ayant des degrés divers de complexité. L'émergence de la vie serait délimitée, d'un côté, par un seuil de complexité génétique et, de l'autre, par un ou plusieurs seuils de complexité biochimique. Au lieu de converger vers un objet unique à l'interface de la biologie et de la physique, les approches réductive et constructive semblent donc au contraire pointer vers les frontières de leurs champs disciplinaires respectifs (Fig. 10).

Cette situation est en partie responsable du flottement terminologique que l'on observe dans la littérature. Certains chercheurs n'hésitent pas considérer leurs créatures de laboratoire comme « vivantes » dès lors

81 La reproduction se fait par division spontanée. L'évolution peut se faire par incorporation de nouvelles molécules venant modifier (catalyser, inhiber, etc.) les réactions chimiques de base.

qu'elles satisfont certains critères plus ou moins arbitraires choisis par eux, sans se rendre compte qu'ils refuseraient probablement cette appellation à d'autres systèmes qui satisfont les mêmes critères. D'autres chercheurs, plus prudents, les appellent parfois des systèmes « suprachimiques » ou « infrabiologiques », ou les considèrent comme des

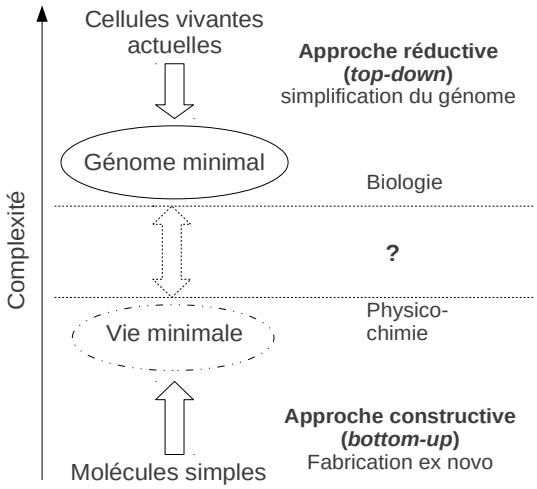

Figure 10: Cellules minimales : convergence ou fossé ?

« approximations à la vie cellulaire » qu'ils disent « similaires à des systèmes vivants », ou « presque mais pas tout à fait vivants », ou « partiellement vivants », ou encore « semi-vivants » comme disait Haldane lui-même. Dans quelle catégorie ranger des entités qui remplissent certaines mais pas toutes les conditions qu'ils jugent nécessaires pour appartenir à la catégorie des êtres vivants ? Ces entités sont capables d'auto-maintien mais pas d'auto-reproduction, ou vice-versa, ou leur auto-reproduction est limitée à quelques générations, ou elle n'est pas limitée mais ne donne pas lieu à une évolution darwinienne, et lorsqu'elles sont capables de se reproduire et d'évoluer, c'est le métabolisme ou la membrane qui leur font défaut. Le problème n'est pas la capacité humaine à créer des formes de vie artificielles, mais la surabondance d'objets artificiels potentiellement vivants que nous sommes en mesure de créer.

L'un des arguments souvent répétés pour expliquer l'absence d'une définition ou d'une théorie de la vie est que nous n'en connaissons jusqu'à maintenant qu'une seule instance, celle qui s'est développée sur Terre, de sorte que nous n'avons rien à quoi la comparer pour en tirer une généralisation. L'argument cesse d'être valable dès que l'on se penche sur les réalisations présentes et futures de la biologie synthétique : la plupart de ces monstres de laboratoire n'ont jamais existé sur Terre et certains sont assez éloignés des formes de vie connues. Nous pourrions

donc commencer à mener dans les laboratoires une étude comparative des êtres vivants comme nous le ferions sur une exoplanète… si seulement nous savions lesquelles parmi ces créatures sont effectivement vivantes.

L'une des promesses de la biologie synthétique était que la création de vie en laboratoire nous aiderait à en comprendre les secrets et à en donner une définition. Dans cette perspective, les critères formulés au départ par les chercheurs (métabolisme, reproduction, évolution, etc.) pouvaient apparaître comme des hypothèses de travail attendant d'être confirmées. Or, nous sommes actuellement dans une situation où, quelle que soit l'hypothèse de départ, nous trouverons sans doute le moyen de créer des entités qui y soient conformes ; et quelles que soient les entités créées, nous pourrons sans doute formuler des définitions de la vie ou des listes de critères qui leur correspondent. C'est-à-dire que nous pourrons avoir une définition pour chaque entité, et une entité pour chaque définition. Les micelles autopoïétiques de Luisi et Varela sont-elles vivantes ? Oui et non, suivant la définition que l'on adopte. Au lieu de trancher parmi les définitions concurrentes, la biologie synthétique multiplie les réponses alternatives.

La situation est différente dans le domaine des origines de la vie, car les hypothèses concurrentes sont confrontées à une réalité historique. L'hypothèse d'un monde à ARN comme étape dans l'histoire de la vie sur Terre est soit vraie soit fausse indépendamment de nos connaissances et de notre capacité à en reproduire les conditions *in vitro*. En revanche, l'hypothèse selon laquelle la vie se caractériserait par un conteneur, un métabolisme et un support informationnel (Bedau, 2007; López García, 2007) a des conditions de vérité beaucoup moins claires.

3. Complexité systémique et minimalité fonctionnelle

Lorsque l'on parle de vie ou de cellules minimales, il convient de distinguer deux questions différentes. La première porte sur la catégorisation et la définition du vivant : *Qu'est-ce que la vie ?* La seconde porte sur sa minimalité : *Quels sont les objets les plus petits ou les plus simples qui appartiennent à cette catégorie ?* L'une des suppositions de l'approche constructive était que la réponse à la seconde question permettrait de répondre à la première, c'est-à-dire que l'exploration des limites et des conditions de possibilité des êtres vivants nous aiderait à comprendre ce qu'est la vie de façon plus générale. Cette supposition pose à son tour deux questions, à savoir : *Quel est le critère de minimalité pertinent pour explorer les frontières de la vie ?* et *Saurons-nous discriminer les objets qui se situent d'un côté et de l'autre de la frontière ?*

Selon l'approche constructive, la vie est une propriété émergente associée à un état d'organisation de la matière. Comme disent Luisi, Chiarabelli et Stano (2006, p. 608) :

> « [I]t is assumed that cellular life on Earth originated from the inanimate matter, via an accretion of molecular and supramolecular complexity, up to the point where structures were produced, that had the novel and emergent property of being 'living' ».

Le seuil de minimalité, depuis cette perspective, devrait correspondre à la complexité systémique minimale requise pour qu'émerge la propriété en question. Cela pose un double problème. D'un côté, la notion de complexité n'est pas définie dans la littérature sur les cellules minimales[82]. De l'autre, nous ne savons pas quelle est la propriété censée émerger lorsque le seuil de complexité est atteint. Est-ce la propriété de se maintenir loin de l'équilibre thermodynamique ? Est-ce la mise en place de réseaux moléculaires auto-catalytiques ? Est-ce la conjonction du métabolisme et des capacités auto-reproductives ? Est-ce l'apparition de mécanismes de sélection naturelle rendant possible une évolution darwinienne ?

Si le processus conduisant à la vie est continu, alors il n'y a peut-être pas de seuil non-arbitraire séparant le vivant de l'inerte. Dans ces conditions, dit Luisi, il est sans doute très difficile de formuler une définition univoque de la vie, mais cela ne veut pas dire que toutes les définitions soient équivalentes :

---

82 Comme le reconnaît Melanie Mitchell (2009, p. 95), du *Santa Fe Institute* consacré à l'étude des systèmes complexes, il n'existe à l'heure actuelle aucune définition universelle de la complexité, car chacune des sciences et des disciplines qui s'y intéressent a sa propre idée, parfois très informelle, de ce que cette notion signifie. Au-delà d'une compréhension intuitive — et subjective — de la complexité, il n'existe donc pas de définition précise commune à la biologie et à la physique, ni même à l'intérieur de chacune de ces disciplines (Adami, 2002). Tandis que la physique des systèmes dynamiques, par exemple, s'intéresse à la complexité des processus, définie à partir de la théorie de l'information, la complexité biologique fait plutôt référence à la forme des organismes, à leurs fonctions, ou aux séquences génétiques qui les codent (Adami, 2002; Edmonds, 1999). Et bien qu'il existe des tentatives de définition de la complexité biologique à partir de la thermodynamique et de la théorie de l'information (Adami, Ofria, & Collier, 2000), elles ne sont pas présentes ni reprises en biologie synthétique. Lorsqu'il traite des phénomènes d'auto-organisation sous contrôle thermodynamique, par exemple, Luisi (2006) parle de complexité sans la définir et de manière intuitive, au sens où les produits d'un processus d'auto-organisation sont plus complexes que leurs constituants. La complexité y est associée à l'ordre résultant de l'auto-organisation, mais le lien entre les deux n'est pas explicitement établi ni scientifiquement justifié. Pour une liste non-exhaustive des différentes mesures de la complexité, voir Seth Lloyd (2001).

> « [I]t is clear that the process leading to life is a continuum process, and this makes an unequivocal definition of life very difficult. […] In view of this arbitrariness on where to put the marker, is any definition equally good? Surely not, as one definition may be more meaningful than another, depending on what you want to do with it. In fact, the following criteria appear important: a definition of life should permit one to discriminate between the living and the non-living in an operationally simple way and it should not be too restrictive […] All forms of life we know about should be covered by such a definition » (Luisi, 1998, p. 616-617)

On ne peut pas exiger des biologistes qu'ils répondent à certaines questions (« *Qu'est-ce que la vie ?* », « *Où se situe la frontière entre le vivant et l'inerte ?* ») avant d'avoir mené les recherches qui permettraient d'y répondre. L'exploration des différents modes d'organisation de la matière situés entre les processus chimiques les plus simples et les organismes vivants actuels pourrait par exemple mettre en évidence une espèce de transition de phase, dont les paramètres restent à établir, entre la matière inanimée et la matière vivante (Fig. 11). Dans ce cas, l'étude détaillée du phénomène nous révélerait ses caractéristiques et ses conditions de possibilité. Nous pourrions alors préciser les notions de complexité et de minimalité en leur donnant une signification physique et, dès lors, proposer une définition opérationnelle de la vie couvrant toutes ses manifestations possibles.

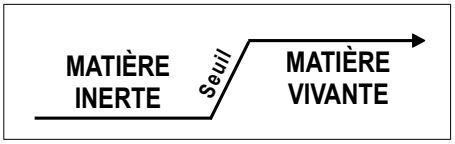

Figure 11: Transition de phase entre deux états d'organisation de la matière sur une ligne de complexité croissante.

Si une telle « transition de phase »[83] venait à être identifiée, cela suffirait-il pour affirmer, comme le faisait Jacques Monod dans les années 1970, que le mystère de la vie est enfin résolu ? Peut-être pas. En revanche, ce qui ne fait aucun doute, c'est que cela nous permettrait d'expliquer un certain nombre de phénomènes caractéristiques des êtres vivants. Cela étant, expliquer et définir sont deux activités différentes, et il se trouve que nous connaissons déjà (voir Schrödinger, 1944) quelque chose de semblable à une telle « transition de phase » avec, d'un côté, des systèmes dont l'entropie interne augmente inexorablement et, de l'autre, des systèmes qui maintiennent ou réduisent leur entropie tout en se situant loin de l'équilibre thermodynamique. Cela nous permet, comme l'exige Luisi dans le passage cité ci-dessus, de discriminer entre le vivant

83 L'idée d'une transition de phase entre la non-vie et la vie est avancée sérieusement par Morowitz et Smith (2007) pour expliquer la raison d'être physico-chimique des êtres vivants. Voir note 88, p. 161.

et l'inerte d'une manière simple et opérationnelle, et cela couvre toutes les formes de vie que nous connaissons[84].

Imaginons par ailleurs que les chercheurs découvrent une « transition de phase » entre deux états d'organisation de la matière dont la frontière ne correspond pas à la distinction intuitive que nous établissons entre le vivant et l'inerte et qu'elle ne couvre donc pas toutes les formes de vie que nous connaissons. (La même chose s'est produite avec la définition moléculaire de l'eau : plusieurs liquides que nous appelions « eau » n'avaient pas la composition chimique H_2O.) Serions-nous prêts à remettre en question notre précompréhension de la vie et nos pratiques classificatoires ou dirions-nous que le phénomène identifié, quoique extrêmement intéressant pour comprendre les êtres vivants, ne saisit pas toute la complexité de ces derniers et que, en fin de compte, le mystère de la vie n'est pas encore levé ? Le simple fait de poser la question montre que la réponse n'est pas évidente et que l'intuition pourrait éventuellement prévaloir[85].

De plus, l'identification d'une telle « transition de phase » entre deux états d'organisation de la matière nous renseignerait certainement sur les limites « inférieures » de la vie mais pas forcément sur ses limites « latérales ». C'est-à-dire que le fait d'identifier un seuil minimal d'organisation de la matière ne nous aiderait peut-être pas à déterminer le statut vital de systèmes complexes — matériels ou pas — qui ne sont pas des systèmes chimiques. Par exemple, si la transition de l'inerte au vivant avait lieu aussitôt que s'enclenche un mécanisme de sélection naturelle de réplicateurs, selon l'hypothèse de Richard Dawkins (1976) actualisée par Addy Pross (2011), devrait-on considérer pour autant que les virus informatiques et les mèmes sont des êtres vivants ou bien les constituants essentiels de systèmes vivants plus larges comme Internet et les cultures humaines ? Autre exemple : si la transition de l'inerte au vivant avait lieu dès qu'un système est suffisamment complexe pour réagir de manière non stéréotypée à des stimuli et s'adapter aux variations de son environnement — autrement dit : pour acquérir une espèce d'autonomie ou de cognition (Bitbol & Luisi, 2004) —, devrait-on considérer pour autant que HAL, l'ordinateur de *2001 : l'odyssée de l'espace* est une créature virtuelle et néanmoins vivante ? Et qu'en est-il de R2D2 et C6PO, les robots mécaniques de *La guerre des étoiles* ? Quels que soient les principes d'organisation qui déclenchent la transition de l'inerte au vivant, nous serons sans doute capables d'implémenter les mêmes principes dans des systèmes matériels non-chimiques ou dans des systèmes virtuels. Ces systèmes potentiels sont-ils vivants ? La question reste posée[86].

84 Les cristaux pourraient également satisfaire la condition thermodynamique.

85 Quoique, comme nous le disions plus haut, l'intuition perd pied dès que l'on s'éloigne des terrains qui lui sont familiers.

L'identification d'un seuil de complexité est apparemment plus simple du côté de l'approche réductive, car la minimalité y est comprise comme le plus petit nombre de gènes nécessaires pour que la cellule accomplisse ses fonctions de base. Pourtant, le critère génétique n'est pas non plus très clair.

S'il s'agit d'identifier la cellule la plus simple, alors le nombre de gènes n'est sans doute pas le critère le plus pertinent. D'abord parce que le séquençage du génome de plusieurs espèces a montré que la complexité d'un organisme n'est pas proportionnelle au nombre de gènes codants[87]. Ensuite, parce que les gènes codants ne représentent qu'une petite partie du matériel génétique fonctionnel de la cellule ; parce qu'il existe aussi des éléments cellulaires qui ne dépendent pas ou peu de ce patrimoine génétique, mais qui sont néanmoins essentiels, comme les organites (mitochondries et chloroplastes) des cellules eucaryotes ; parce que d'autres fonctions importantes, voire essentielles, de l'organisme dépendent d'éléments « étrangers » à la cellule, comme les plasmides des bactéries, les endosymbiotes des protozoaires ou les microbiotes des mammifères (flore intestinale, par exemple) ; et enfin parce que la complexité d'un système n'est pas déterminée par le nombre de ses composants mais par les relations qu'ils entretiennent (Lenski, Ofria, Collier, & Adami, 1999). En effet, un système complexe est d'autant plus petit que ses éléments sont mieux intégrés. Or, un organisme est un système intégré et pas une collection d'éléments isolés. L'analyse des relations entre tous les éléments de la cellule devrait donc a priori constituer un critère de complexité plus fiable que le simple dénombrement d'une partie des éléments du génome.

Par ailleurs, la présence d'un gène dans l'ensemble minimal que veulent créer les chercheurs est déterminée par sa fonction puisqu'un gène ne peut être éliminé que si la fonction qu'il remplit n'est pas essentielle pour le fonctionnement du système dans son ensemble. Or, le nombre de fonctions n'est pas égal au nombre de gènes (un même gène ou une même protéine pouvant remplir plusieurs fonctions) et dépend en grande partie de l'intégration du système. Sachant par ailleurs que la plupart des définitions du vivant sont formulées en termes de propriétés

86 Répondre par la négative, en insistant sur la nature chimique de la vie, comme dans la définition de Joyce (1994), ne serait-ce pas établir une distinction similaire à celle des vitalistes entre matière vivante et inanimée ?

87 Bien que le ver *C. elegans* ne mesure qu'un millimètre de long et ne compte qu'un millier de cellules, son génome contient environ 20 000 gènes codant pour des protéines, c'est-à-dire 50 % de plus que la drosophile et presque autant que l'humain (21 000). La tomate possède quant à elle 31 760 gènes, soit 50 % de plus que nous (Wade, 2012). Si la complexité était proportionnelle au nombre de gènes codants, il faudrait admettre que les tomates sont plus complexes que les humains.

fonctionnelles, alors c'est le nombre de fonctions et pas le nombre de gènes qui devrait être pertinent pour décrire une cellule vivante minimale, même si en pratique il est beaucoup plus simple de compter le nombre de gènes — ou le nombre de parties d'un organisme — que le nombre de fonctions (McShea, 2000).

Quelle que soit l'approche adoptée pour créer une cellule vivante minimale, la minimalité en question doit d'abord être comprise en termes de complexité fonctionnelle. C'est-à-dire que l'on devrait théoriquement commencer par établir le schéma fonctionnel minimal d'une cellule vivante pour ensuite chercher l'ensemble minimal de composants (gènes, réactions physico-chimiques, etc.) qui permettent d'implémenter ces fonctions. Chacune des propriétés essentielles du vivant (par exemple : métabolisme, auto-reproduction, évolution) est réalisée par un système fonctionnel composé de sous-systèmes composés à leur tour de sous-systèmes moins complexes. La fonction de chacun de ces sous-systèmes est leur contribution au fonctionnement du système organique dans son ensemble. Bien entendu, l'un des problèmes de cette approche — au-delà des difficultés techniques — est l'identification des propriétés fonctionnelles de base à partir desquelles commencer l'analyse[88].

Un autre problème est la signification du concept de fonction. La littérature en biologie synthétique l'emploie un peu à la manière de Cummins (1975), c'est-à-dire comme synonyme de capacité, mais elle ne précise pas si les capacités en question doivent être effectivement réalisées ou seulement réalisables (bien qu'elles n'aient pas de métabolisme, les spores bactériennes sont capables de le réactiver lorsque les conditions s'y prêtent). De plus, une capacité jugée essentielle peut faire défaut à certains individus ou types d'individus sans que l'attribution de vie leur soit pour autant retirée, laissant supposer une conception normative à la manière de Wright (1973). Les mules, par exemple, ne peuvent pas se reproduire ni évoluer ; ce sont pourtant des êtres vivants dont le système reproductif est considéré comme dysfonctionnel. Si le critère de minima-

88 C'est ce que font Harold Morowitz et Eric Smith (2007) en étudiant la raison d'être des êtres vivants d'un point de vue de physique et en analysant les mécanismes universels qui les rendent possibles : de même que les éclairs et les ouragans, les êtres vivants sont selon eux des canaux (*channels*) permettant de transporter, bien plus rapidement que ne le ferait la simple diffusion, des flux de matière et d'énergie entre deux réservoirs de potentiels différents. Tandis que l'éclair est un canal de transport de potentiel électrique, la vie est un canal de transport dans le domaine chimique. Partant de cette hypothèse, les auteurs analysent les réseaux de réactions métaboliques et anaboliques qui rendent ce transport possible et identifient dans le cycle réductif de l'acide citrique un réseau autocatalytique fondamental sur lequel tout le noyau du réseau anabolique pourrait s'appuyer. Cela permet ensuite à Morowitz et Srivinasan (2009b, 2009a) de proposer le schéma chimique minimal et universel d'un « chémoautotrophe réductif ».

lité doit être conçu en termes fonctionnels, le flou du concept s'ajoute à la difficulté de la mesure.

Par ailleurs, la minimalité fonctionnelle est relative à un environnement. Il est bien évident qu'une plante verte n'est capable de réaliser la photosynthèse qu'en présence de lumière et d'eau. Plus précisément, elle ne peut exercer cette capacité que dans des conditions données. Dans le noir ou en l'absence d'eau, elle en exercera d'autres. C'est-à-dire que les diverses capacités des plantes vertes et des êtres vivants en général leur permettent de réagir et de s'adapter à des conditions environnementales variées. Or, une façon de réduire le génome d'une cellule consiste à éliminer les fonctions qui rendent cette adaptation possible. Cela consiste à mutiler un organisme fonctionnel en le privant de ce qui fait de lui un système autonome dans son environnement naturel et à suppléer aux fonctions perdues en le plaçant dans un environnement artificiellement contrôlé. En effet, plus le milieu extérieur est constant et plus il est facile pour un système organique d'assurer l'homéostasie de son milieu intérieur. Ainsi, les bactéries synthétiques de Craig Venter ne sont viables que dans un milieu de culture très riche, ce qui fait dire à un éditorialiste de *Nature* qu'il s'agit en fait de « vie sous perfusion » :

> « That last sounds like a detail, but is in fact essential. The minimal requirements depend on the environment — on what the organism does and doesn't have to synthesize, for example, and what stresses it experiences. Too much minimization and you end up with cells on life support. » (Nature, 2007)

Toutes les capacités qu'une cellule minimale n'est plus capable de réaliser doivent être prises en charge par le milieu. Cela n'est pas sans rappeler, comme le suggère la citation précédente, la situation des malades en salle de réanimation qui ne survivent que dans la mesure où certaines de leurs fonctions vitales, comme la respiration, sont assurées de l'extérieur par des moyens techniques. La situation n'est cependant pas tout à fait la même puisque les malades gardent la possibilité de récupérer certaines de leurs fonctions perdues tandis que les cellules minimales dont nous parlons sont constitutivement dépendantes pour leur survie d'un « maintien » externe.

Cela pose la question des limites de la vie depuis une perspective nouvelle. En effet, si l'on peut externaliser les fonctions de base d'une cellule vivante, alors on peut aussi concevoir que les virus soient des cellules vivantes ayant externalisé leurs fonctions métaboliques et leurs fonctions de transcription vers les cellules hôtes, à la manière de parasites extrêmes. La découverte en 2011 de *Megavirus chilensis* vient appuyer cette hypothèse[89]. Cela signifierait que les frontières de la vie minimale se

89 Les virus géants à ADN actuels pourraient être issus d'un ancêtre cellulaire commun ayant progressivement perdu une partie de son génome après être

situent bien en deçà de ce que la plupart des biologistes seraient prêts à admettre.

D'un autre côté, si l'on tient compte du fait que tous les êtres vivants sont plus ou moins dépendants d'un milieu et d'un écosystème, lequel prend en charge un certain nombre de capacités qu'autrement ils devraient assurer par eux-mêmes, alors la complexité systémique minimale nécessaire à la vie ne se limite pas à celle des individus pris isolément mais doit inclure celle des écosystèmes dont ils font partie. Dans la mesure où elles requièrent une culture humaine technologiquement avancée pour les maintenir en vie, les cellules de Craig Venter n'ont donc rien de simple et sont au contraire terriblement complexes.

4. Vie, fonctions et sélection naturelle

4.1. Fonctions et sélection

Toutes les capacités d'un système ne sont pas équivalentes et la littérature sur les cellules minimales s'intéresse exclusivement à celles qui d'une manière ou d'une autre sont « utiles » aux êtres vivants. De même que le cœur est capable de produire des pulsations sonores à intervalles réguliers sans que cela ait un impact sur le fonctionnement de l'organisme qui l'abrite, les cellules et leurs composants ont des capacités qui ne leur sont d'aucune utilité ou qui au contraire sont potentiellement délétères. La recherche de minimalité ne concerne donc que les capacités que l'on estime « positives » d'une manière ou d'une autre. C'est-à-dire que les fonctions qu'on leur attribue ne sont pas comprises au sens de

---

devenu le parasite obligatoire d'autres organismes cellulaires (Arslan, Legendre, Seltzer, Abergel, & Claverie, 2011; Legendre, Arslan, Abergel, & Claverie, 2012). Des expériences menées sur *Mimivirus* montrent qu'après 150 générations dans un environnement contrôlé (à l'intérieur d'amibes), la réduction de son génome atteint 16 % et provoque des modifications morphologiques et fonctionnelles (Boyer et al., 2011). Même s'ils ne sont pas des organismes autonomes, le statut vital des virus est comparable, selon certains auteurs, à celui des cellules d'un organisme pluricellulaire (Hegde, Maddur, Kaveri, & Bayry, 2009). Par ailleurs, la découverte en 2013 de *Pandoravirus* suggère l'existence d'un quatrième domaine du vivant, car ce nouveau type de virus géant est quasiment sans homologie avec les trois domaines reconnus : eucaryotes, eubactéries et archébactéries (Philippe et al., 2013; Yong, 2013). Cette découverte, affirme le communiqué du CEA, comble une discontinuité entre le monde viral et le monde cellulaire qui a été érigée en dogme depuis les fondements de la virologie moderne dans les années 1950. Plus récemment, la découverte en 2014 d'une nouvelle famille de virus géants, *Pithovirus*, conforte encore cette hypothèse (Legendre et al., 2014).

Cummins (1975) ou de Davies (2001) comme étant n'importe quelle capacité systémique suffisamment complexe pour requérir une analyse fonctionnelle, mais plutôt comme celles qui contribuent positivement (et pas accidentellement) à… la vie. Or, si l'on ne sait pas *a priori* ce qu'est la vie on peut difficilement savoir ce qui y contribue au-delà des réponses dictées par l'intuition et la familiarité.

Une façon de résoudre ce problème est de réduire progressivement les capacités d'une cellule vivante jusqu'à observer dans le système un changement qualitatif manifestement incompatible avec la vie (quoi que cela signifie). C'est la méthode suivie par l'approche réductive à travers la simplification du génome. Elle ne distingue pas les gènes « utiles » des gènes « inutiles » mais ceux qui sont essentiels (collectivement) de ceux qui ne le sont pas. Une autre façon de résoudre le même problème consiste à laisser l'évaluation des capacités à la sélection naturelle : si une capacité donnée a été sélectionnée, c'est sans doute parce qu'elle est — ou a été — utile. Cette seconde méthode correspond à une conception différente de la vie, non plus systémique mais darwinienne, c'est-à-dire historique. Selon Bedau (1996, 1998, 2007), elle présente l'avantage de fournir une justification aux caractéristiques fonctionnelles qui chez d'autres auteurs apparaissent comme des listes de critères hétérogènes. Si la vie repose avant tout sur une évolution par sélection naturelle, les diverses fonctions des êtres vivants sont seulement des moyens de cette fin et la minimalité fonctionnelle est le nombre minimal de fonctions nécessaires à sa réalisation.

Tandis que la conception systémique du vivant recherche une propriété émergente associée à un seuil de complexité de la matière, la conception darwinienne désigne l'évolution par sélection naturelle comme principe d'organisation tendant vers une complexification indéfinie des objets qui lui sont soumis. Quoi que soit la vie et quelles que soient les caractéristiques des systèmes vivants existant sur Terre ou ailleurs, l'évolution par sélection naturelle est le mécanisme causal permettant d'y arriver. Comme dit Cairns-Smith : « *Without natural selection the whole adventure would never have got off the ground. That kind of in-built ingenuity that we call 'life' is easily placed in the context of evolution:* life is a product of evolution. »

Les inconvénients de la conception darwinienne de la vie sont analogues à ceux rencontrés par l'approche historique des fonctions. Imaginons que l'on observe sur Mars une entité apparemment identique à l'un de nos lapins terrestres. On pourrait difficilement dire, en partant d'une définition invoquant le critère d'évolution darwinienne (Bedau, 1998; Joyce, 1994), s'il s'agit ou pas d'un être vivant. Il faudrait en effet analyser sa constitution jusqu'au niveau moléculaire, chercher des traces fossiles de ses ancêtres potentiels et/ou attendre pour voir s'il se reproduit et si ses descendants évoluent. « *Il est intéressant de noter*, disent Maynard Smith & Szathmary (1999, p. 7), *que l'observation d'un objet ne permet*

pas de déterminer s'il est vivant ou s'il est le produit de l'activité d'êtres vivants. Seule la connaissance de l'histoire de l'objet permet de faire cette distinction. »

Les trois principales objections formulées à l'encontre de l'approche historique des fonctions sont donc applicables à cette conception de la vie.

1. De même que l'on peut attribuer une fonction à un organe sans avoir la moindre idée de son étiologie, on peut reconnaître un être vivant sans connaître sa généalogie ni analyser sa constitution chimique.
2. Les définitions fondées sur la sélection naturelle reposent sur des bases contingentes, car si la théorie darwinienne s'avérait fausse, cela voudrait dire qu'il n'y a pas d'êtres vivants ni de fonctions biologiques.
3. Il semble contre-intuitif qu'un organisme soit privé de vie parce qu'il n'a pas d'histoire. D'autant plus contre-intuitif que la biologie synthétique cherche justement à créer des êtres vivants qui, à l'instar des doubles accidentels, n'ont ni ancêtres ni descendants.

4.2. Les limites de la sélection

Pour comprendre l'origine de ces objections, il faut distinguer entre les définitions de la vie et du vivant. Tandis que les premières se rapportent au phénomène général, les secondes s'appliquent aux entités individuelles. La conception darwinienne n'est valable que pour des populations et sur plusieurs générations (Luisi, 1998, p. 618). Pour savoir si un individu est vivant conformément à cette conception, il faudrait par exemple le rattacher à ce que Millikan (1984) appelle une famille reproductivement établie : si le lapin martien appartient à une espèce de lapins dont on peut remonter l'arbre phylogénétique pour constater son évolution, alors on dira que le lapin est vivant dans la mesure où la famille à laquelle il appartient satisfait la définition. Une autre solution consiste à mesurer indirectement l'adaptation d'une population à un milieu en observant par exemple, comme dans l'une des expériences menées sur Mars par la sonde Viking, la quantité de gaz (déchets organiques ?) produite par un échantillon de sol mélangé à une solution dite « nutritive » : si après une période de stabilité la production de gaz augmente, on pourra supposer que l'activité chimique est le fait d'agents capables non seulement de se reproduire mais aussi de s'adapter (Chao, 2000). Un modèle informatique a été développé par Bedau et Packard (1992) pour démontrer la possibilité d'une mesure de l'évolution d'un système et proposer un test de vie extraterrestre. Le paradoxe est que ces tests pourraient nous apprendre s'il y a de la vie sur Mars, mais pas si notre lapin martien est un être vivant. Comme disent Bedau et Packard (1992, p.

31), il n'y a vraiment que la biosphère tout entière — le réseau complexe d'organismes en interaction — dont on puisse dire qu'elle est vivante.

Toute définition qui implique une évolution darwinienne est problématique dans la mesure où elle repose sur un état futur ou passé du système et sur d'autres systèmes (Fleischaker, 1990). Il existe cependant une autre solution consistant à adopter une interprétation dispositionnelle comparable à celles que Bigelow et Pargetter (1987) et Walsh (1996) formulent à propos des fonctions. Lorsque la NASA définit la vie comme étant un système chimique auto-maintenu capable d'évolution darwinienne, on peut entendre cette capacité non pas d'un point de vue historique, comme un fait avéré, mais plutôt comme une disposition des individus liée à leurs caractéristiques présentes. Si le feu n'est pas vivant selon la définition de la NASA, c'est parce qu'il ne possède aucun mécanisme de transmission héréditaire de ses éventuelles variations. En biologie synthétique, l'interprétation dispositionnelle correspond aux tentatives d'intégrer des molécules de support d'information dans des vésicules auto-reproductives (Forster & Church, 2006; Luisi, Ferri, et al., 2006). La présence de mécanismes d'auto-reproduction, de variation et d'hérédité y apparaissent en effet comme les conditions minimales de possibilité d'un être vivant, indépendamment de son histoire passée et de sa descendance future.

L'interprétation dispositionnelle répond aux trois objections précédentes, mais on peut lui en opposer d'autres. Imaginons une exoplanète où les conditions environnementales sont tellement constantes à long terme que les créatures qui y vivent, parfaitement adaptées, n'ont plus besoin d'évoluer[90, 91]. Ces créatures se trouvent par ailleurs à l'abri de

90 Même si l'environnement abiotique restait parfaitement constant, de sorte que les variations internes d'une espèce adaptée à son milieu n'apportent pas de valeur adaptative supplémentaire susceptible d'être sélectionnée, il existerait néanmoins des facteurs d'évolution biotiques comme la course aux armements entre espèces (théorie de la reine rouge).

91 À titre d'exemple, les stromatolites sont des bio-structures apparues il y a près de 3,5 milliards d'années qui continuent d'être produites actuellement, et, parmi les cyanobactéries responsables de leur construction, les eontophysalis sont le plus long lignage vivant connu sur la planète. Cela ne veut pas dire que les bactéries en question n'ont pas évolué. Plus près de nous, les cœlacanthes sont apparus il y a 350 millions d'années (MA) et n'ont pratiquement pas changé morphologiquement jusqu'à aujourd'hui. La famille des limules est apparue il y a plus de 500 MA et l'une de ses espèces actuelles (Limulus polyphemus) existe depuis 220 MA sans modifications notables (« Horseshoe crab », 2010). Les myxines modernes ressemblent beaucoup à un ancêtre fossile (Myxinikela siroka) daté de 300 MA (« Introduction to the Myxini », 2010). Et parmi les requins, apparus il y a plus de 400 MA, certaines espèces ont adopté leur forme moderne il y a presque 100 MA (« Origin of Modern Sharks », 2010). Pour comparaison, les premiers animaux

nombreuses causes externes de mortalité. Après tout, les bactéries et les animaux vivant à plusieurs kilomètres de profondeur dans la roche (Borgonie et al., 2011 ; Drake, 2011), sous la glace (Loveland-Curtze, Miteva, & Brenchley, 2009), ainsi que les habitants des sources chaudes du fond océanique, se trouvent dans des conditions relativement similaires. Imaginons aussi que, à l'instar de nombreux organismes terrestres, elles disposent de mécanismes d'auto-réparation les protégeant du vieillissement[92]. Ces créatures seraient virtuellement immortelles et n'auraient plus besoin de se reproduire[93]. Or, à partir du moment où la reproduction aurait pour elles un coût tout en n'apportant aucun bénéfice, on pourrait imaginer qu'elles perdissent leurs capacités reproductives et ne fussent alors plus sujettes à l'évolution par sélection naturelle. Pour autant, cesseraient-elles d'être vivantes ?

À la suite de Lovelock (1995; 1974) on peut aussi concevoir que l'unité fondamentale du vivant soit la biosphère tout entière (Gaïa), dont les organismes individuels feraient partie à la manière des cellules d'un organisme plus grand (A. Goldman, 2011; Margulis & Sagan, 1995; Rowe, 1992, 1998). Dans cette optique, la Terre elle-même serait un être vivant, incapable de se reproduire[94] mais capable de métabolisme, d'homéostasie — ou plutôt d'homéorhésie — et d'évolution non

---

sont apparus vers 610 MA, les plantes terrestres autour de 450 MA, et les dinosaures il y a 230 MA. Cela dit, la stase morphologique n'implique pas une stase évolutive, car les organismes continuent d'être soumis à une pression sélective pour s'adapter aux variations de leur environnement. Par ailleurs, le taux d'évolution des organismes est variable : chez les bactéries, le taux de mutation optimal est d'autant plus faible que l'environnement est constant (Radman, 2002); chez les tortues de mer, on a observé un taux d'évolution très lent (Avise, Bowen, Lamb, Meylan, & Bermingham, 1992; FitzSimmons, Moritz, & Moore, 1995); et au niveau macroévolutif, selon la théorie des équilibres ponctués, l'évolution des espèces est marquée par de longues périodes de stase (évolution lente ou stagnation) ponctuées de rares et brèves périodes de modification et de spéciation rapides, mais ces variations du taux d'évolution ne sont pas forcément liées à l'environnement.

92 Des bactéries sont capables de vivre pendant plus d'un demi million d'années avec une activité métabolique cellulaire assurant la réparation de leur ADN (S. S. Johnson et al., 2007).

93 D'après Klarsfeld et Revah (2000), la reproduction est le moyen le plus simple de répondre à la fois au processus interne du vieillissement et aux causes externes de destruction. Les mécanismes d'auto-réparation peuvent pallier au vieillissement et permettre à des arbres de survivre pendant des dizaines de siècles, mais ils ont un coût et ils ne peuvent pas empêcher la foudre, les incendies, etc. Le meilleur moyen de les éviter est la reproduction. Par ailleurs, c'est aussi le meilleur moyen de s'adapter (grâce au mécanisme de sélection naturelle). Mais si un organisme n'a plus besoin de s'adapter et s'il est à l'abri à la fois de la dégradation et de la destruction, a quoi peut lui servir la reproduction ?

darwinienne. La sélection naturelle y jouerait un rôle essentiel pour l'émergence de comportements complexes d'auto-régulation à l'échelle planétaire tout en n'agissant que localement et en complément d'autres mécanismes géophysiques (Lenton, 1998). Si la vie n'est pas un accident unique, on peut imaginer que les galaxies fourmillent d'écosphères vivantes sortant du cadre de la théorie darwinienne. Au-delà de l'image poétique, l'intérêt de cette idée est de proposer une conception compréhensive du vivant, et non plus seulement réductrice et analytique, où la vie des organismes individuels (au sens courant) ne dépend pas de leur constitution systémique (organisation interne) ou historique (évolution par sélection naturelle) mais de leur contribution à un système plus grand (relation parties–tout). Or, c'est aussi sur la base d'une relation de type parties–tout que nous allons définir le concept de fonction au CHAP. VII.

Une conception basée sur la sélection naturelle, qu'elle soit tournée vers le passé ou le futur, rencontre les mêmes problèmes d'identification de ses frontières que la conception systémique, avec d'un côté un super-organisme planétaire et, de l'autre, des cristaux d'argile, des molécules auto-réplicantes et des virus. D'après l'hypothèse de Cairns-Smith (1986), les cristaux d'argile font de bons candidats à l'origine de la vie et pourraient eux-mêmes représenter une forme de vie préorganique dans une perspective darwinienne[95]. En effet, ils croissent, s'auto-répliquent, sont sujets à évolution par sélection naturelle et leurs surfaces catalysent la formation de molécules organiques qui pourraient être les précurseurs de la vie telle que nous la connaissons. L'idée a été reprise par Bedau (1991) pour critiquer les insuffisances des définitions des fonctions biologiques basées sur la sélection naturelle. Or, tout l'argument de Bedau repose sur l'évidence (non justifiée) selon laquelle les cristaux d'argile ne sont pas vivants, raison pour laquelle nous serions réticents à leur attribuer des fonctions.

---

94 À moins que l'humanité parvienne à terraformer et coloniser d'autres planètes. Et à moins que des morceaux de roche terrestres expulsés dans l'espace par une éruption volcanique ou un impact de météorite puissent « ensemencer » d'autres planètes en y transportant des bactéries et autres microbes.

95 « That somewhat ill-defined collection of strange attributes that now we can so easily recognize as 'life' would have emerged only slowly as explicitly the *product* of evolution through natural selection. And that process itself would have depended on the prior existence of evolvable entities. These would have been 'replicators' or 'primary organisms' made, in the view to be proposed, largely out of clay minerals. These first evolvers, whatever their material basis, would have belonged to some class of physicochemical system too simple to be yet 'alive', yet capable of evolving under natural selection so that 'life' could sooner or later emerge. » (Cairns-Smith & Hartman, 1986, p. 16)

Nous n'hésitons pourtant pas à attribuer des fonctions à certaines molécules organiques bien que nous ne les reconnaissions pas comme vivantes. Et nous ne les reconnaissons pas comme telles malgré leur conformité à la définition de Joyce (1994). L'évolution darwinienne est en effet présente au niveau moléculaire dans l'hypothèse d'un monde à ARN, voire à une étape antérieure (Gogarten, 2011; Pross, 2011), et si elle constitue la marque du commencement de la vie alors on devrait accepter la possibilité d'une vie moléculaire, c'est-à-dire purement chimique, non seulement dans le lointain passé de la Terre (Joyce, 1989, 2002), mais aussi au fond des éprouvettes des laboratoires de biologie synthétique.

L'exemple des virus est encore plus frappant. Considérés jusqu'à récemment comme de simples boîtes à génome inertes, nous savons désormais que les virus peuvent être aussi grands et presque aussi complexes que certaines bactéries et cellules eucaryotes[96]. Probables contemporains du monde à ARN, ils pourraient être les « inventeurs » de l'ADN ou du noyau cellulaire et ont certainement contribué de manière essentielle aux mécanismes d'évolution de la vie depuis ses origines jusqu'à nos jours[97]. Sujets à une évolution par sélection darwinienne,

96 Avec une taille de 400 nm et un génome de 1,2 millions de paires de bases (pour plus d'un millier de gènes), *Mimivirus* et *Megavirus chilensis* sont plus grands que de nombreuses bactéries et se distinguent des autres virus connus par la possession de gènes codant pour des fonctions cellulaires comme la réparation de l'ADN ou la traduction et la synthèse protéiques (Arslan et al., 2011; La Scola et al., 2003; Raoult et al., 2004). *Mimivirus* se caractérise également par le fait de pouvoir être à son tour infecté par d'autres virus plus petits appartenant à une nouvelle famille d'entités biologiques appelées virophages (La Scola et al., 2008). Avec une taille avoisinant un micromètre de long et près de 2,5 millions de paires de bases codant pour plus de 1100 gènes, *Pandoravirus salinus* et *Pandoravirus dulcis* sont deux fois plus grands que *Megavirus* et n'ont quasiment aucun point commun avec les virus géants précédemment caractérisés (Philippe et al., 2013). Ils sont dépourvus de capside et seulement 6% des protéines codées par leur génome ressemblent à des protéines déjà répertoriées dans les autres virus ou les organismes cellulaires. Quant à *Pithovirus*, découvert en 2014, il atteint une taille exceptionnelle de 1,5 microns de long et 0,5 microns de diamètre (Legendre et al., 2014).

97 Bien que l'origine des virus soit toujours sujette à controverses, la découverte récente de *Mimivirus*, *Mamavirus*, *Marseillevirus, Cafeteria roenbergensis* virus (CroV), *Megavirus chilensis*, *Pandoravirus* et *Pithovirus* plaide en faveur de l'ancienneté des virus dans l'arbre du vivant, avec un ancêtre cellulaire antérieur à LUCA, le dernier ancêtre commun de tous les organismes cellulaires actuels (Arslan et al., 2011; Forterre, 2006; E. C. Holmes, 2011). Mais les mécanismes de transferts de gènes horizontaux rendent difficile l'identification de cet ancêtre commun. D'autres auteurs attribuent quant à eux une origine commune aux cellules eucaryotes et aux virus géants, créant pour ces derniers une quatrième branche de l'arbre du vivant (Claverie,

habitants permanents de nos cellules et contributeurs de notre patrimoine génétique, les virus brouillent la frontière entre le vivant et l'inerte :

> [L]e parasitisme obligatoire que l'on croyait être l'apanage des virus existe aussi chez les bactéries, des organismes indéniablement vivants ; des virus contiennent de l'ARN et de l'ADN, alors que l'on pensait caractériser ces micro-organismes par la présence d'un seul acide nucléique ; enfin, certains virus changent de forme en dehors de tout contexte cellulaire, ce que l'on tenait pour impossible. Ces découvertes confèrent aux virus un rôle inédit, celui de précurseurs de la vie, et en font même un moyen de définir ce qu'est la vie [c'est-à-dire ce qui peut être infecté par un virus]. (Saïb, 2006, p. 61)

Au problème des types d'objets satisfaisant les conditions de la conception darwinienne s'ajoute celui des frontières temporelles de la sélection. Si la vie d'une entité dépend de sa capacité *avérée* à évoluer par sélection naturelle, le nombre de générations de descendants nécessaires pour que l'évolution se produise dépendra évidemment des entités elles-mêmes et de leur contexte environnemental. Il ne sera donc pas forcément le même pour deux populations d'une même entité. On se souvient que dans une approche historique, la fonction d'un trait est ce pour quoi il a été sélectionné, de sorte que les premières générations de ce trait sont dénuées de fonction (voir l'exemple des pseudo-lions de Neander, p. 66). Il n'y a pas de transfert à rebours de la propriété fonctionnelle acquise par sélection. Si l'on interprète de la même manière l'acquisition de la vie, alors les premières générations d'un type d'entité ne peuvent pas être considérées vivantes même si leurs descendants le sont. Cela ne pose pas de problème tant qu'on se limite à la vie « naturelle », car toutes les entités vivantes actuelles ont hérité cette propriété d'un ancêtre commun (cellulaire, moléculaire, minéral…) perdu dans le lointain passé de la Terre. Le problème se pose avec les organismes synthétiques, car non seulement il serait difficile de dire quelle est la première génération d'individus vivants dans une population donnée, mais deux populations d'une même entité dans des environnements différents pourraient ne pas avoir le même statut, l'une étant vivante et l'autre pas, alors que rien ou presque ne les distingue[98].

2005). Pour un débat en faveur et contre l'inclusion des virus dans l'arbre de la vie, voir respectivement (Forterre, 2010; Raoult & Forterre, 2008; Villarreal & Witzany, 2010) et (Moreira & Lopez-Garcia, 2009).

98 Une interprétation dispositionnelle dirait plutôt que l'évolution darwinienne de l'une des populations confirme que cette capacité était présente dès le départ dans les deux populations.

4.3. Frontière temporelle et pertinence

Selon Godfrey-Smith, la frontière temporelle des fonctions biologiques est une question de pertinence :

> « Some might wonder how recent the selective episodes relevant to functional status have to be. The answer is not in terms of a fixed time—a week, or a thousand years. Relevance fades. Episodes of selection become increasingly irrelevant to an assignment of functions at some time, the further away we get. » (Godfrey-Smith, 1994, p. 356)

Pour connaître la fonction d'un trait, il faut tenir compte des épisodes sélectifs dont il a fait l'objet. Or, plus on remonte loin dans le passé de ce trait et moins les épisodes de sélection seront pertinents pour justifier une attribution fonctionnelle. De même, pour comprendre le mobile et les circonstances d'un crime, il n'est généralement pas utile de remonter jusqu'à l'enfance des personnes impliquées, et encore moins jusqu'à celle de leurs parents. La même idée est valable pour l'attribution de vie à des populations d'organismes synthétiques dans une perspective darwinienne.

On comprend bien que la pertinence à laquelle fait référence le passage ci-dessus est une question purement épistémique, comme lorsque Cummins parle de la pertinence explicative de l'analyse fonctionnelle. Il s'agit de savoir jusqu'où remonter dans le passé pour justifier l'attribution de fonctions (ou de vie) à un trait actuel. La question se complique quand on cherche à savoir jusqu'où il est possible de faire remonter cette attribution. Autrement dit, si un trait x possède actuellement la fonction f (ou si un type de système est actuellement considéré comme vivant), jusqu'où peut-on faire remonter la possession de f (ou l'appartenance à la catégorie des êtres vivants) ? Ou encore : si on remonte suffisamment loin dans le passé, à partir de quel moment pourra-t-on affirmer que le trait x a acquis la fonction f (ou que le système est devenu vivant) ? La pertinence, ici, ne concerne plus seulement la justification des attributions mais les attributions elles-mêmes. Elle acquiert de ce fait une portée ontologique : à quel moment un système devient-il vivant ; à quel moment un trait acquiert-il une nouvelle fonction ?

En réinterprétant le passage ci-dessus, on dira que l'attribution de vie et de fonctions perd de sa pertinence au fur et à mesure que l'on remonte dans le temps. Trois cas de figure se présentent. Dans le premier, il s'agit d'un problème épistémique : notre connaissance des processus de sélection à l'œuvre dans une population est d'autant plus limitée que la fenêtre temporelle d'observation est courte. En remontant dans le passé, on ne fait que perdre de la perspective. À l'instant t la population tout entière ou certains de ses membres pourraient donc avoir acquis une

nouvelle fonction ou être devenus vivants sans que nous ayons eu le temps de nous en apercevoir.

Dans le second cas de figure, ce n'est pas notre connaissance qui est en question mais le système lui-même qui, à l'instant *t*, ne possède pas forcément un état vital et fonctionnel bien défini : ou bien parce que la transition d'un état à un autre prend un certain temps, de sorte que la frontière entre les deux est floue ; ou bien parce que ces propriétés sont intrinsèquement diffuses, de sorte que le système n'est pas soit vivant soit inerte mais à la fois l'un et l'autre à des degrés différents. L'idée d'une frontière temporellement floue entre deux états bien distincts est en accord avec l'intuition et avec l'usage courant des concepts de vie et de fonction : un système est vivant ou il ne l'est pas, un trait possède une fonction ou il ne la possède pas, bien qu'on admette l'existence d'effets de bord. Le problème est qu'une évolution très lente peut transformer une période d'indétermination transitoire en un état quasi-permanent[99].

Il peut sembler plus raisonnable d'adopter, pour les processus évolutifs continus comme la spéciation, l'acquisition ou la perte de fonctions et l'apparition de la vie, une ontologie graduelle où la possession d'une propriété et l'appartenance à une catégorie obéissent à une logique paraconsistante (Bruylands, Bartik, & Reisse, 2010; Vásconez & Peña, 1996). Dans une telle ontologie, le même système peut être à la fois vivant et non-vivant (ou fonctionnel et non-fonctionnel) et l'attribution de vie (ou de fonctions) n'être que partiellement vraie. En dépit des apparences, cette ontologie présente l'avantage d'être mieux conforme à la manière naturelle que nous avons de diviser en catégories le monde qui nous entoure (voir CHAP. V, SECT. 6.1).

Dans le troisième cas de figure, ce n'est pas la connaissance ni l'ontologie qui sont en question mais le langage. L'identification des frontières du vivant et des fonctions pose problème dans la mesure où l'on essaie de fixer une limite quantitative à des notions qualitatives. On peut considérer par exemple, en reprenant l'exemple du paradoxe sorites, que ce qui distingue un tas d'un non-tas ce n'est ni une propriété particulière ni le nombre de grains qui le composent mais le qualificatif qu'on lui applique pour des raisons de communication, car il permet d'objectiver quelque chose qui en soi ne constitue pas un objet[100]. Plus on lui enlève de grains et moins il est pertinent — en termes communicatifs — de le décrire comme un tas. À partir de quel moment cesse-t-il de l'être ? Peut-être à partir du moment où les interlocuteurs cessent de se comprendre. Les limites d'un terme vague sont les limites de la communication : elles ne sont pas définies par avance et dépendent à la fois du

99 Voir note 91, p. 166.

100 Pour un aperçu de la « *vagueness* » et du paradoxe sorites, que nous ne voulons pas discuter ici, se reporter aux entrées correspondantes de l'Encyclopédie Stanford de Philosophie (Hyde, 2008; Sorensen, 2008).

contexte et de l'usage, c'est-à-dire des échanges entre les membres de la communauté linguistique. De ce point de vue, la vie et les catégories fonctionnelles serviraient à désigner des objets que l'on veut rassembler (pour une raison ou pour une autre) sans que les critères d'appartenance aux ensembles ainsi formés aient été préalablement fixés. L'extension de ces concepts est l'objet d'une négociation au sein de la communauté scientifique et elle est sujette à révisions[101]. Au fur et à mesure que nous découvrirons ou fabriquerons des objets que certains estimeront pertinent d'inclure dans la catégorie des êtres vivants de nouvelles frontières devront être tracées et les définitions révisées. Il n'y a pas de réponse préétablie. Et comme l'a montré le débat sur les fonctions, les réponses des uns et des autres ne sont pas les mêmes selon les disciplines qu'ils pratiquent.

Le problème de la pertinence des épisodes de sélection pour l'attribution de vie et de fonctions n'est pas sans rappeler celui rencontré par les différentes équipes de physiciens impliqués dans la découverte des particules W^+, W^- et Z^0 porteuses de l'interaction nucléaire faible. Comme le rapporte Peter Galison dans *How Experiments End* (1987), les chercheurs appartenant à des « écoles » différentes ne donnaient pas la même interprétation et n'attribuaient pas la même importance aux données expérimentales recueillies concernant la question de savoir si ces particules avaient été observées ou pas. La pertinence respective des preuves statistiques et des preuves visuelles pour justifier la détection des particules était en effet relative aux pratiques de recherche des différents groupes d'investigateurs ; ce qui était considéré comme suffisant pour les uns ne l'était pas pour les autres et vice-versa. En l'absence d'une norme préétablie, la décision de conclure les expériences a donc été le résultat d'une négociation entre les parties. L'existence des bosons n'est pas négociable[102], mais la justification expérimentale de cette existence l'est bel et bien, et il n'existe pas de réponse définitive à la question de savoir quelles données empiriques prouvent cette existence ni si un « cliché de colli-

101 La polémique suscitée par le résultat positif de l'une des expériences biologiques du programme Viking visant à détecter de la vie sur Mars (DiGregorio, 2000) illustre le type de discussion qui pourrait avoir lieu face à une éventuelle forme de vie extraterrestre ou laborantine. Certains diront que la chose est vivante, d'autres défendront le contraire, des analyses complémentaires seront réalisées, et lorsqu'un consensus finira par émerger la décision sera prise. Quelle qu'elle soit, cette décision renforcera l'une des définitions alternatives : celle qui inclut (ou exclut) la chose en question.

102 Du moins tant que l'on se situe à l'intérieur du Modèle Standard et que l'on ne prend pas trop au sérieux l'ontologie de la physique quantique, car on peut « choisir » d'adopter une autre interprétation rendant compte des mêmes observations sans recourir à l'existence de particules (à ce sujet, voir par exemple « La crise de l'atomisme contemporain » dans Bitbol, 1998, p. 5).

sion » constitue ou pas une preuve de détection. Par analogie, si l'existence des êtres vivants n'est pas négociable, il n'existe peut-être pas de réponse définitive à la question de savoir quels épisodes de sélection justifient l'attribution de vie et de fonctions à un individu ou à une population.[103]

103 Pour une critique différente de l'argument de Griffiths sur la pertinence temporelle, voir Karen Neander (2010).

Chapitre V

La vie existe-t-elle ?

1. Faut-il définir la vie ?

En février 2017, trois planètes de taille similaire à la Terre étaient découvertes dans la zone d'habitabilité de l'étoile Trappist-1, située à 40 années–lumière de distance (Gillon et al., 2017). Quelques mois plus tôt, l'exoplanète la plus proche de nous, Proxima Centauri b, était découverte à une distance de « seulement » 4,2 années–lumière, orbitant elle-aussi dans la zone habitable de son étoile et dotée d'une masse comparable à celle de la Terre (Anglada-Escudé et al., 2016). Depuis 1995, plus de 3 400 exoplanètes ont été confirmées, dont plus de 350 de type terrestre et ayant une température de surface compatible avec la présence d'eau liquide (NASA, 2017).

Ces découvertes ont alimenté les recherches et les discussions en exobiologie ; une science née dans les années soixante avec les premiers projets d'exploration spatiale (Lederberg, 1960). Moins de dix ans après la conquête de la Lune, les sondes Viking effectuaient déjà, en 1976, des tests pour détecter des traces de vie microbienne dans le sol martien. En 2020, la mission européenne ExoMars déposera sur cette planète une plateforme et un véhicule équipés d'instruments scientifiques avec un but similaire. D'autres missions en cours ou en projet, comme Europa Lander, s'intéressent aux lunes de Jupiter et de Saturne à la recherche de traces de vie extraterrestre, présente ou passée. Les objets que ces missions recherchent sur d'autres mondes ressemblent à ceux que nous connaissons déjà sur Terre, notamment chez les extrêmophiles (Cavicchioli, 2002; Trent, 2007; Westall et al., 2015), et les biosignatures qu'elles cherchent à détecter sont font partie des caractéristiques de la vie terrestre, comme la chiralité des acides aminés (Creamer, Mora, & Willis, 2017; ESA, 2017). Cependant, l'un des problèmes de l'exobiologie est que les formes de vie nouvelles que nous découvrirons peut-être ailleurs

ne ressemblent pas forcément à celle que nous connaissons déjà ici. Saurons-nous les reconnaître si elles existent ? C'est la question que beaucoup se posent.

Jusqu'à présent, la reconnaissance du vivant, ou la discrimination entre l'animé et l'inerte, était une tâche en grande partie intuitive et innée, liée chez l'humain comme chez d'autres animaux à des besoins écologiques (identification rapide des proies et des prédateurs). Cette capacité intuitive était ensuite complétée chez nous par l'apprentissage et l'évolution des connaissances, de sorte que — hormis quelques cas limites et peu familiers — il était relativement facile pour les biologistes et pour les profanes de se mettre d'accord sur le statut vital d'un objet donné. Aujourd'hui, la conquête spatiale rend les choses beaucoup plus difficiles, car elle ouvre la possibilité de découvrir des objets qui se situent peut-être au-delà des limites de notre intuition et de nos connaissances actuelles.

La biologie synthétique et les recherches sur la vie minimale posent un problème similaire. Nous avons acquis très récemment la capacité de transformer et de créer des objets qui partagent certaines des caractéristiques du vivant tel que nous le connaissons. Mais comment distinguer parmi tous ces objets ceux qui sont vivants de ceux qui ne le sont pas ? Par exemple, les « cristaux vivants » créés par des biophysiciens le sont-ils vraiment (Palacci, Sacanna, Steinberg, Pine, & Chaikin, 2013) ? Peut-être pas, reconnaissent les auteurs, mais ils soulèvent le problème de la frontière entre ce qui est actif et ce qui est vivant. De manière générale, les produits de la biologie synthétique tendent à rendre plus floues les frontières entre les organismes et les machines (Deplazes & Huppenbauer, 2009) et à remettre en question des notions qui nous semblaient jusqu'à présent intuitives.

D'une part, il est évident que ces recherches contribuent au progrès des connaissances concernant les phénomènes eux-mêmes. Ainsi, comme disent Alonso Ricardo et Jack Szostak (2009), la capacité de créer de la vie *ex novo* nous aidera sans doute à comprendre comment la vie peut commencer, quelle est la probabilité qu'elle existe aussi sur d'autres mondes et, en définitive, ce qu'est la vie. D'autre part, ces recherches alimentent le débat sur la signification des notions de « vie » et de « vivant ». André Brack rapporte à ce propos une anecdote significative :

> À l'occasion du colloque sur la vie tenu à Modène en 2000, les organisateurs ont demandé à chacun des membres de la Société internationale pour l'étude sur l'origine de la vie de donner une définition de la vie. Des soixante-dix-huit réponses reçues, il fut impossible de dégager une réponse commune. (2007, p. 15)

Ces 78 réponses (Palyi, Zucchi, & Caglioti, 2002) n'épuisent pas l'éventail des définitions possibles. Edward Trifonov (2011, 2012) en analyse 123 pour essayer de leur trouver un dénominateur commun, et

Radu Popa (2004) en recueille quant à lui plus de 300 dans la littérature scientifique.

Cette profusion définitionnelle est sans doute un phénomène assez récent et limité à certains domaines de recherche, car pendant longtemps les biologistes n'ont pas eu besoin de définir la vie[104] (Popa, 2010). Linus Pauling disait à ce propos qu'il est parfois plus facile d'étudier un objet que le définir. De son côté, J.B.S. Haldane commençait son ouvrage *Qu'est-ce que la vie ?* (1949) en affirmant qu'il ne répondrait pas à cette question et qu'il doutait que l'on pusse jamais lui donner une réponse complète. Et lors d'une conférence portant sur la définition de la vie, François Jacob déclara : « *Cette question n'a pas de réponse* » (2000). D'après Pier Luigi Luisi, la situation n'est pas différente parmi les chercheurs en biologie synthétique :

> Ces chercheurs devraient savoir ce qu'ils recherchent ou ce qu'ils essaient de reproduire dans leurs laboratoires. Ce n'est pas le cas : les définitions actuelles de la vie sont rares […], et celles qui existent ne sont pas très populaires. (1998, p. 613)

L'une des raisons de cette impopularité pourrait être celle pointée par Jack Szostak (2012) lorsqu'il affirme que les tentatives de définition de la vie sont arbitraires, illusoires, et ne nous aident pas à en comprendre les origines. Ou bien celle de Claude Bernard qui dans ses *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux* disait qu'« [i]l n'y a pas de définition des choses que l'esprit n'a pas créées et qu'il n'enferme pas tout entières; il n'y a pas, en un mot, de définition des choses naturelles ». Ou tout simplement le fait que de nombreux concepts scientifiques, comme celui de « continent » en géologie, n'ont pas de définition précise, car celle-ci n'est pas nécessaire.

Cependant, la multiplication récente des publications et des colloques qui s'interrogent sur la définition de la vie exprime un besoin de la part des chercheurs dans les domaines de l'exobiologie, de la biologie synthétique, de la vie artificielle, et des origines de la vie (Bersini & Reisse, 2007). D'une part, il s'agit pour eux de se mettre d'accord sur une définition opératoire pour des raisons de communication (Mix, 2015). De l'autre, il s'agit d'essayer de comprendre la vie elle-même, de façon universelle et pas seulement sous les formes concrètes qui nous sont jusqu'ici familières (Damiano & Luisi, 2010; Popa, 2010; Ruiz-Mirazo, Peretó, & Moreno, 2010).

---

104 On remarquera cependant que le nombre d'apparitions de l'expression « definition of life » dans le corpus d'ouvrages indexés par Google connaît un pic significatif autour de 1880 (entre 1860 et 1930), puis se maintient relativement stable entre 1940 et 2008 à un niveau trois fois moins élevé (voir Google Ngram https://goo.gl/SgDhzB).

Les définitions en question se présentent généralement sous la forme d'une liste de critères ou de propriétés communes à tous les êtres vivants (connus), comme celle de Mayr (1997), ou sur la base d'un modèle qui décrit le fonctionnement de l'objet que l'on veut définir, comme le chemoton de Tibor Gánti (2003), ou encore sur la base d'une théorie générale, comme la théorie autopoïétique. Cela étant, comme le signale Jean Gayon, le terme « vie » n'intervient guère en biologie contemporaine comme un terme théorique entrant dans des hypothèses ou modèles, car contrairement à d'autres concepts abstraits qui ne sont pas non plus bien définis dans d'autres sciences (par exemple « énergie », « force »), la « vie » ne désigne pas une entité non observable intervenant dans des hypothèses fondamentales visant à expliquer certaines classes de phénomènes. Pourtant, ajoute cet auteur, les définitions sont des constructions théoriques qui s'appuient idéalement sur une ou plusieurs charactéristiques que l'on croit être essentielles à la chose définie (Gayon, 2010a, p. 238).

Connaître ou reconnaître l'essence de quelque chose — ou que quelque chose a une essence — n'implique pas forcément que l'on sache la définir. C'est à peu près ce que disait Haldane en 1949 pour justifier que l'on ne puisse pas répondre entièrement à la question-titre de son ouvrage : nous savons ce que cela fait que d'être vivant (*what it feels like to be alive*), de même que nous savons ce que sont le rouge, la douleur ou l'effort, mais nous n'arrivons pas les décrire entièrement en termes d'autres choses. C'est aussi ce que disait, depuis une autre perspective, l'Agence Spatiale Européenne sur son site web :

> « Life is a fascinating thing. We can all recognise it but we cannot define it. It is the most elusive natural property known to science and no definition adequately captures its essence. » (ESA, 2008)

D'après Cleland et Chyba (2002), les tentatives de définition de la vie se trouvent face à un dilemme analogue à celui rencontré par ceux qui prétendaient définir l'eau avant l'existence de la théorie moléculaire[105]. Ce qui nous manque, disent-ils, ce n'est pas une définition mais une théorie générale de la vie qui permettrait d'expliquer l'ensemble de ses propriétés à partir de sa nature profonde et en laissant de côté le sens ordinaire du mot. On se souvient que Millikan défendait aussi à propos des fonctions l'idée d'une définition théorique analogue à la définition moléculaire de l'eau, et Bedau dit à peu près la même chose à propos de la vie :

105 Searle (1992, p. 101) défendait une idée similaire à propos du « mystère » de la conscience, laissant entendre que le « mystère » de la vie avait été résolu par la biologie moléculaire. Selon lui, une connaissance adéquate de la conscience au niveau neuro-physiologique lèverait le mystère.

> « We want to know what life is, not what people think life is. Glass does not fall under the everyday concept of a liquid, even though chemists tell us that glass really is a liquid. Likewise, we should not object if the true nature of life happens to have some initially counterintuitive consequences. » (Bedau, 1996)

Depuis cette perspective, le terme « vie » ne désigne pas une entité non observable, comme la force vitale, mais plutôt un genre naturel, comme l'eau, ou un phénomène naturel, comme l'électricité, que l'on peut identifier sous différentes formes dans différents objets. Il n'entre pas dans l'explication théorique du comportement des objets que l'on dit « vivants », mais désigne plutôt le comportement particulier qui fait qu'on les distingue des autres objets naturels. Car au-delà du problème de la définition de la vie, la biologie est traversée par une tension historique entre deux idées centrales. D'un côté, l'idée qu'il existe une distinction naturelle entre les êtres vivants et les objets inanimés. De l'autre, l'idée qu'il existe une continuité ontologique fondamentale entre les deux domaines, laquelle a été démontrée par la biologie moléculaire. Ces deux idées correspondent respectivement à l'affirmation et à la négation d'une spécificité de la vie par rapport à la matière inanimée.

Par exemple, après avoir dressé la liste des caractéristiques du vivant tel que nous le connaissons, Ernst Mayr insistait sur la distinction entre le vivant et l'inerte en affirmant qu'elle est au fondement de l'autonomie de la biologie en tant que science :

> Toutes ces caractéristiques des organismes vivants les distinguent fondamentalement des systèmes inanimés. Dans l'histoire des sciences, c'est la reconnaissance graduelle du caractère unique et distinct des êtres vivants qui a conduit à délimiter cette science qu'on appelle la biologie, et à reconnaître son autonomie. (Mayr, 1997, p. 35)

Dans le contexte de l'exploration spatiale des années 1970, Richard Dawkins s'interrogeait pour sa part sur ce qui est commun à toutes les formes de vie que l'on pourrait éventuellement découvrir ailleurs dans l'univers (sa réponse étant l'évolution par sélection naturelle) :

> « When astronauts voyage to distant planets and look for life, they can expect to find cratures too strange and unearthly for us to imagine. But is there anything that must be true of all life, whereever it is found, and whatever the basis of its chemistry? If forms of life exist exist whose chemistry is based on silicon rather than carbon, or ammonia rather than water, if creatures are discovered that boil to death at 100 degrees centigrade, if a form of life is found that is not based on chemistry at all but on electronic reverberating circuits, will there still be any general principle that is true of all life? » (Dawkins, 1976, p. 191-2)

En analysant les citations précédentes, on peut y distinguer au moins trois idées indépendantes, mais pas inconnexes. D'abord, l'idée que la vie est un phénomène, une propriété ou un genre naturel dont il faudrait formuler la théorie. Ensuite, l'idée que tous les êtres vivants, quelles que soient les formes qu'ils puissent adopter, ont quelque chose en commun. Finalement, l'idée selon laquelle les êtres vivants se distinguent fondamentalement des objets inanimés. Cette dernière idée semble impliquer l'existence d'une frontière naturelle non-ambigüe entre les deux domaines, de telle sorte qu'un objet donné soit clairement vivant ou non vivant. Cependant, certains estiment qu'un même objet peut être les deux choses à la fois, c'est-à-dire que son appartenance au domaine du vivant est une question de degrés (Bruylands et al., 2010). D'autres ajoutent que la vie n'est pas seulement une question de degrés, mais aussi de types, c'est-à-dire qu'il y aurait plusieurs façons d'être vivant réclamant chacune d'elles une définition différente (Malaterre, 2010).

Il est possible que les êtres vivants ne soient pas une classe naturelle (Keller, 2007; voir aussi Morange, 2007, p. 65) et que la distinction entre les êtres vivants et non-vivants ne le soit pas non plus. S'il en était ainsi, quelles conséquences cela pourrait-il avoir pour la biologie en tant que science de la nature ? Pour essayer de répondre à cette question, nous allons nous pencher sur le problème de la définition du concept de planète en astronomie.

2. Les êtres vivants sont-ils un genre naturel ?

2.1. À propos de la définition de « planète »

Pendant des siècles, il fut facile de distinguer et de classer les objets astronomiques. Outre le Soleil et la Lune, le ciel comprenait les astres fixes de la voûte céleste (les étoiles), les astres errants (les planètes) et les phénomènes transitoires (comètes, météores). Après Copernic et Galilée, la Terre elle-même fut classée parmi les planètes et la Lune fut déclassée. De 1781 à 1950, de nouvelles planètes furent découvertes, dont Uranus, Cérès, Pallas, Juno, Vesta, Astrae, Neptune et plusieurs autres dont le nombre ne cessait d'augmenter. En 1851, William Hershel proposa de classer les corps les plus petits dans une nouvelle catégorie, les astéroïdes, de sorte que le nombre de planètes se réduisit à huit[106]. Pluton ne fut

106 Face à la multiplication du nombre de petites planètes, on décida de faire précéder leur nom du numéro correspondant à l'ordre de leur découverte : 1 Cérès, 2 Palas, 3 Junon, etc., créant ainsi un système de classification qui perdure jusqu'à maintenant. La qualité des instruments de l'époque ne permettant pas de résoudre leur disque, on les appela « astéroïdes » (comme des étoiles) ou « planètes mineures », les distinguant ainsi des autres planètes.

découverte qu'en 1930 et demeura pendant plus de soixante ans une planète solitaire aux confins du système solaire. De même que le statut de Cérès fut remis en question après la découverte d'une vaste population d'objets avec des masses et des orbites similaires, celui de Pluton ne fut reconsidéré que lorsque les astronomes découvrirent dans les années 1990 une vaste population d'objets comparables dans la Ceinture de Kuiper. Or, puisque certains d'entre eux, comme Eris (d'abord nommée Xena ou 2003 UB_{313}), sont aussi grands que Pluton, ils devraient eux-aussi être des planètes ; et s'ils ne le sont pas, alors Pluton ne devrait pas l'être non plus. À la même époque, la découverte des planètes extrasolaires, des planètes vagabondes et des naines brunes allait compliquer davantage la situation.

Bien que les planètes aient toujours été au cœur de leur discipline, les astronomes ne se penchèrent vraiment sur la définition scientifique de cette notion que lorsque leur identification devint problématique. En 2005, l'Union Astronomique Internationale (UAI) mis en place une commission spéciale à cet effet. Après plus d'un an de travail, elle proposa une définition qui fut votée et approuvée par l'Assemblée générale du 24 août 2006 (IAU, 2006). Une planète y est définie comme :

> Un corps céleste qui (a) est en orbite autour du Soleil, (b) possède une masse suffisante pour que sa gravité l'emporte sur les forces de cohésion du corps solide et le maintienne en équilibre hydrostatique, sous une forme presque sphérique, et (c) a éliminé tout corps susceptible de se déplacer au voisinage de son orbite. »

Étant donné que Pluton ne remplit pas la troisième condition, la proposition de l'UAI incluait la création d'une nouvelle catégorie de corps célestes, les « planètes naines » (qui ne sont pas des planètes), à laquelle Pluton, Cérès, Eris et d'autres étaient reléguées[107]. Depuis ce jour, le nombre officiel de planètes est diminué à huit.

Un astronome ayant pris part à la commission chargée de formuler la proposition déclara peu après que le but était de trouver une base scientifique pour la nouvelle définition et qu'ils avaient choisi la gravité comme étant le facteur déterminant, ajoutant que « c'est la Nature qui décide si un objet est une planète ou pas » (Britt, 2006). Un autre spécialiste justifiait la décision en affirmant que la catégorie des planètes n'est pas arbitraire mais objective, et qu'elle reflète notre compréhension de l'architecture des systèmes planétaires en général :

> « This is not just a debate about words. The question is an important one scientifically. The new definition of a planet reflects advances in our understanding of the architecture of our solar system and others. […] In short, "planet" is not an arbitrary category but an objetive class of celestial bodies. » (Soter, 2007)

107 Depuis juin 2008, les planètes naines en orbite autour du Soleil à une distance supérieure à celle de Neptune sont appelées des « plutoïdes ».

On pourrait interpréter ces paroles en disant que les planètes sont une classe de corps célestes naturellement distincts des planètes naines, des satellites, des comètes, des astéroïdes, etc., et que la définition de l'UAI est un reflet des connaissances scientifiques actuelles qui permet de saisir tous les objets qui sont effectivement des planètes et aucun de ceux qui ne le sont pas. Dans cette perspective, on pourrait également dire que si Pluton a été exclue de la classe des planètes, c'est parce qu'elle n'en fait objectivement pas partie. Les astronomes auraient donc eu tort en 1930 de la considérer comme telle et n'auraient reconnu leur erreur que soixante-seize ans plus tard. Certains estiment même que la distinction entre planètes et non-planètes est quantifiable en théorie et en pratique (Margot, 2015; Soter, 2007). Ainsi, suivant le troisième critère de la définition, les « vraies planètes » seraient au moins 5 000 fois plus massives que la somme des masses de tous les objets au voisinage de leur orbite, tandis que Pluton et les autres « planètes naines » font au contraire partie d'une population d'objets de masses comparables. Cela ne veut pas dire que la définition soit parfaite ni qu'elle ne doive pas évoluer avec le progrès des connaissances, mais il semble qu'elle apporte une solution au problème de l'identification des planètes que les découvertes récentes avaient posé.

En examinant plus en détail les circonstances qui entourent cette définition, on s'aperçoit que les choses ne sont pas aussi simples. La proposition originale de l'UAI, publiée le 16 août, élevait à douze le nombre de planètes, en y incluant Cérès (précédemment considérée comme un astéroïde), Charon (précédemment considéré comme une lune de Pluton) et 2003 UB_{313} (Eris, Xena). Or, Michael Brown, le découvreur de cette dernière et le principal bénéficiaire de la proposition, fit campagne contre elle auprès de ses collègues, car il considérait qu'elle était incohérente et arbitraire (M. E. Brown, 2010). D'après lui, les options scientifiquement acceptables impliquaient soit l'exclusion de Pluton de la classe des planètes, soit l'inclusion de dizaines ou de centaines de corps comparables, mais beaucoup d'astronomes s'opposaient à l'une ou à l'autre, de sorte que la proposition de l'UAI était un compromis destiné à ne fâcher personne. L'absence de consensus et les critiques formulées par lui et par d'autres contre cette première définition conduisirent à l'élaboration d'une nouvelle proposition au milieu d'une forte controverse.

Aujourd'hui, la controverse continue. En 2017, des astronomes de la NASA qui refusent la définition de l'UAI proposent de la remplacer par une définition géophysique qui restituerait à Pluton son statut de planète et qui en élèverait le nombre à près de 110 (Runyon et al., 2017). Ils défendent la seconde condition (quasi-rotondité) et rejettent les autres. La première implique que la définition de l'UAI n'est valable que pour *notre* Système Solaire et laisse en suspend le statut des objets orbitant d'autres étoiles ou errant librement dans la galaxie. La troisième, disent-

ils, n'est remplie par aucune des planètes de notre Système Solaire, car de petits corps sont constamment injectés à l'intérieur des orbites planétaires. De plus, elle est relative à la masse et à la distance, de sorte que même un objet de la taille de la Terre situé dans la Ceinture de Kuiper ne serait pas en mesure de nettoyer son orbite et ne serait donc pas considéré comme une planète. En ce qui concerne la seconde condition, elle est critiquée par d'autres pour son imprécision, car la quasi-rotondité est non seulement relative à la composition chimique de l'objet, mais aussi et surtout sujette à une décision arbitraire (Flatow, 2006; « IAU definition of planet », 2010).

Ce genre de débats n'a absolument rien de surprenant, car il fait partie du fonctionnement normal de l'activité scientifique. Ce qui est plus surprenant, en revanche, c'est que plusieurs des participants reconnaissent volontiers que leurs arguments ne sont pas strictement scientifiques, mais aussi de nature sentimentale, culturelle, pédagogique, économique ou politique.

Alan Stern est l'un des principaux détracteurs de la définition de l'UAI. C'est aussi le responsable de la mission *New Horizons*, dont la sonde avait été envoyée vers Pluton six mois avant son déclassement. Le public, dit-il, a du mal a comprendre l'intérêt d'une mission scientifique vers un objet lointain qui n'est pas une planète. Sa définition géophysique vise donc principalement à corriger ce défaut pour que les politiciens continuent de financer l'exploration du Système Solaire (Runyon et al., 2017). Philip Metzger, de la NASA lui-aussi, donne plusieurs raisons scientifiques pour rejeter la définition officielle et retenir le statut planétaire de Pluton, mais il n'hésite pas à donner également des raisons pédagogiques et culturelles (2015).

De leur côté, les partisans de la définition disent en sa faveur qu'elle maintient le nombre de planètes suffisamment bas pour que les élèves puissent les apprendre par cœur (Hackett, 2015). Pour eux, il ne s'agit pas tant de savoir combien de planètes *il y a* dans le Système Solaire, mais de savoir combien nous *voulons* qu'il y en ait, et cette décision n'est pas scientifique mais culturelle. Le président de la commission spéciale, Owen Gingerich, a lui-même affirmé en 2014, lors d'un débat public, qu'il regrettait rétrospectivement que l'UAI ait voulu définir le mot « planète », car c'est un terme défini par la culture dont le sens n'a cessé de changer au cours de l'histoire (ScienceAlert, 2014). Michael Brown, le découvreur de 2003 UB_{313}, fut l'un des premiers à dire que le terme « planète » est culturel et qu'il n'a pas besoin d'une définition scientifique (2010). Son argument repose sur le fait que la définition et le nombre de planètes n'ont en réalité aucune importance astronomique, car quelle que soit celle que l'on adopte, cela ne changera rien au travail scientifique :

> « Consider it this way: if the word *planet* is suddenly redefined to mean either 8 or 53, how will it affect astronomy? Not one tiny bit whatsoever. Astronomers like me will continue to go to telescopes

> and study these objects to learn where they came from and what they are made out of whether they are called "planets", "Kuiper belt objects", or "batholiths." For astronomers, this argument is purely semantic. Who is affected, then? I would argue that it is the public, it is our culture, that would be affected, and, in fact, this is why this is the one astronomical argument, out of the many many many that are out there, that anyone actually seems to care about. In light of this realization, perhaps it makes sense to have a cultural definition of the word planet, rather than a scientific definition. By "cultural definition" what I mean is "what people mean when they say the word planet." » (M. E. Brown, 2006)

Selon cette approche, l'appartenance de Pluton à la classe des planètes n'est pas une question empirique susceptible d'être découverte ni démontrée, car elle dépend avant tout du *choix* d'une définition qui n'est pas elle-même scientifique.

Cela ne veut pas dire que la classification des corps célestes soit arbitraire ni qu'elle soit futile d'un point de vue scientifique. Bien au contraire. Il existe différentes manières de les classer qui sont à la fois pertinentes et importantes. Basri et Brown (2006) ont identifié trois terrains scientifiques sur lesquels il est possible de tracer des frontières.

Le premier est celui des *caractéristiques physiques intrinsèques*. Le facteur déterminant ici est la masse, car dans le continuum quantitatif allant des plus petits astéroïdes aux plus grandes étoiles, les transitions qualitatives sont déterminées par la gravité propre de l'objet. L'une de ces transitions s'effectue lorsque la force de gravité propre d'un corps devient supérieure à la résistance des matériaux qui le composent et le force a adopter une forme sphérique (ou presque). On peut donc distinguer les corps ronds de ceux qui ne le sont pas et tracer entre eux une frontière à la fois précise et pertinente, fondée sur l'équilibre hydrostatique. D'autres transitions importantes se produisent à mesure que la masse du corps s'accroît davantage, par exemple lorsque son énergie gravitationnelle devient suffisante pour modifier la composition chimique des matériaux originels puis pour enclencher un processus de convection. On peut ainsi distinguer les objets qui sont géophysiquement actifs (comme Mars) de ceux qui sont inertes (comme la Lune). En ce qui concerne les géantes gazeuses, d'autres transitions pertinentes se produisent pour des masses supérieures à celle de Jupiter.

Le second « terrain » est celui des *circonstances*. Les astronomes distinguent un satellite d'une planète suivant l'objet autour duquel ils orbitent. Ils distinguent aussi les objets isolés sur leur zone orbitale de ceux qui la partagent. Ainsi, Cérès et Pluton furent considérés des planètes jusqu'à la découverte de corps comparables dans leurs voisinages respectifs. La justification de cette distinction repose sur la dynamique orbitale : certains corps ont une influence gravitationnelle qui leur permet d'éliminer les planétésimaux se trouvant à proximité. Cela

dépend de la masse de l'objet, mais aussi de sa distance à l'étoile centrale ainsi que de la masse et du rayon de celle-ci. Dans notre Système Solaire, le seuil de masse nécessaire ne serait atteint que par Jupiter, tandis que les autres planètes n'ont pu nettoyer leurs orbites que grâce à son influence. Par ailleurs, si la Terre s'était formée ou avait migré dans la Ceinture de Kuiper, elle n'aurait pas été capable de nettoyer son orbite et ne serait dès lors pas une planète selon la définition de l'UAI. Un autre critère pertinent est la stabilité de l'orbite, mais en considérant une échelle de temps suffisante toutes les orbites s'avèrent instables. Si une planète comme Saturne était éjectée un jour du Système Solaire, devrait-on cesser de la classer parmi les planètes ? L'histoire d'un objet ne devrait-il pas être pris en compte ?

Le troisième « terrain » est celui de la *cosmogonie*, c'est-à-dire de l'histoire. Beaucoup d'astronomes pensent par exemple que les planètes sont « par définition » des objets qui se forment dans un disque de gaz autour d'une étoile, par accrétion de planétésimaux, et que c'est là leur caractéristique principale. Depuis cette perspective, si Jupiter n'avait pas de noyau et si elle s'était formée spontanément à partir d'une instabilité gravitationnelle dans la nébuleuse protoplanétaire, alors ce ne serait pas vraiment une planète, parce qu'elle n'a pas l'histoire qu'il faut. Ce qui compte ici, ce ne sont pas les caractéristiques intrinsèques, ni le contexte, mais le mécanisme de formation.

Les astronomes ont donc à leur disposition plusieurs options qui sont à la fois pertinentes et intéressantes d'un point de vue scientifique pour catégoriser leurs objets d'étude à partir de différences qualitatives mesurables. En fonction de leurs intérêts respectifs, ils accorderont sans doute plus d'importance à certains aspects qu'à d'autres. Par exemple, la distinction entre corps actifs et inertes est très intéressante pour les géophysiciens et les planétologues, mais peut-être pas autant pour ceux qui étudient la dynamique et la formation du Système Solaire. Les uns et les autres peuvent donc classer différemment les mêmes corps et les rassembler ou les distinguer pour répondre à des besoins scientifiques différents. Ce travail de catégorisation n'est pas arbitraire, mais il n'est pas non plus dicté strictement par la nature, car il dépend d'un choix relatif à des besoins et des intérêts épistémiques humains.

À laquelle (ou auxquelles) de ces catégories appartiennent les planètes ? À aucune et à toutes. On peut *choisir* de donner ce nom aux objets ronds, ou à ceux qui sont géologiquement actifs, ou à ceux qui ont nettoyé leur orbite, ou bien à ceux formés par accrétion de planétoïdes, ou encore à une combinaison des précédents, mais on ne peut pas dire que les planètes *soient* ceci ou cela, car comme dit Mike Brown, « planète » est une espèce de mot magique qui ne correspond à rien d'un point de vue scientifique :

> « All of the important science of categorization is now done, and done correctly. All that is left to decide is who now gets to use the magical word "planet." There is absolutely no *scientific* argument that anyone could possibly make to prefer one over the other. That would be akin to asking which one is correct. The answer is that they are both correct, and both useful. » (M. E. Brown, 2008)

Dire qu'un genre est *naturel*, c'est affirmer qu'il correspond à un regroupement qui reflète la structure du monde naturel plutôt que les intérêts et les actions des êtres humains (Bird & Tobin, 2017). Nos très lointains ancêtres ont rassemblé sous le nom de planètes les objets célestes qui ne se déplaçaient pas à l'unisson de la voute étoilée. Ce regroupement reposait sur l'observation d'une différence qualitative dans les mouvements célestes et il reflétait la structure du monde naturel puisque les planètes sont effectivement distinctes à tous points de vue des étoiles « fixes ». Plus tard, les astronomes ont rassemblé sous le nom de planètes les objets en orbite autour du Soleil, et sous le nom de satellites ceux en orbite autour d'une planète. Ce regroupement s'appuyait sur une meilleure compréhension de la nature et de la dynamique des objets astronomiques connus et il reflétait encore la structure du monde naturel. Toutefois, leur nombre restait inférieur à dix et leurs caractéristiques étaient relativement similaires. Aujourd'hui, alors que leur nombre a augmenté exponentiellement et que nous connaissons beaucoup mieux la structure du monde naturel, nous ne pouvons plus prétendre que les planètes soient un genre naturel, car le nombre de regroupements possibles a été démultiplié.

2.2. Analogie avec la vie et le vivant

Si, comme le recommandait un éditorial de la revue *Nature* à propos de la biologie synthétique, il serait bon d'admettre que la vie n'est pas un concept scientifique précis[108], alors on pourrait aussi admettre que les êtres vivants ne sont peut-être pas un genre naturel. À l'heure où nous nous préparons à découvrir la vie sur d'autres mondes, à la fabriquer en laboratoire et à la programmer sur ordinateur, la catégorie autour de laquelle Lamarck et Treviranus ont inventé la biologie n'est peut-être plus aussi pertinente qu'il y a deux siècles.

Les biologistes actuels sont capables de faire des distinctions et des classifications de plus en plus fines fondées sur une compréhension plus profonde des objets biologiques et de leur histoire. Récemment, suite à la découverte des virus géants, plusieurs spécialistes ont proposé de créer pour ces derniers une nouvelle branche du vivant située à la base même

108 « It would be a service to more than synthetic biology if we might now be permitted to dismiss the idea that life is a precise scientific concept » (Nature, 2007)

de l'arbre (Raoult & Forterre, 2008; Saïb, 2006). Comme pour Pluton, la question n'est peut-être pas tant de savoir si les virus sont vivants ou pas, mais si nous voulons les considérer comme tels :

> « The question, "are viruses alive?" is typically a philosophical question, meaning that it is our choice to decide if viruses are living entities or not. For a growing number of evolutionists and virologists, viruses should definitively be considered as living entities since they exhibit all features typical of terrestrial life » (Forterre, 2010)

Il n'y a pas une unique manière « correcte » ou scientifiquement pertinente de classer les êtres vivants, ni une essence réelle des groupes biologiques, pas plus qu'il n'existe une essence ni une unique manière correcte de classer les objets astronomiques. La question est de savoir à laquelle de ces classes d'objets nous allons assigner le mot magique « vie ».

Toutes les manières possibles de classer des objets scientifiques ne sont pas équivalentes. Par exemple, la classification phylogénétique est moins arbitraire que la classification traditionnelle et elle est sans doute meilleure que d'autres systèmes classificatoires (Ridley, 1998). De plus, elle admet une unique solution vraie correspondant aux innovations évolutives et aux relations de parenté réelles entre espèces. On peut donc supposer qu'il y a une réponse préétablie à la question de savoir si deux entités appartiennent au même taxon, de même qu'il y a une réponse préétablie concernant l'éventuel lien de parenté entre les virus et les bactéries, les archées et les eucaryotes. Dans le cas où un tel lien serait vérifié, nous n'hésiterions peut-être pas à les considérer comme vivants. Pourtant, l'existence avérée d'une relation de parenté n'est pas nécessaire ni suffisante pour classer les virus parmi les êtres vivants. Elle n'est pas suffisante parce que les relations de parenté s'étendent peut-être en deçà de l'apparition de la vie[109]. Et elle n'est pas nécessaire parce que l'on ne peut pas exclure la possibilité que la vie soit apparue plusieurs fois de manière indépendante[110]. De fait, la vie que nous pourrions découvrir sur

---

109 Dans une perspective évolutionniste, les organismes vivants descendent d'autres entités qui, comme dit Bedau, n'étaient pas encore vivantes mais sur le chemin de la vie (B. Holmes, 2005, p. 4). Les virus pourraient descendre des mêmes entités mais n'avoir pas franchi le pas (ou alors seulement en partie avec les Girus). À l'inverse, des entités autrefois vivantes peuvent avoir des descendants qui ne le sont plus. C'est le cas des endosymbiontes : si les mitochondries descendent de bactéries ayant fusionné avec d'autres plus grandes pour donner lieu aux cellules eucaryotes actuelles, ces entités ne sont plus elles-mêmes vivantes (puisque faisant partie d'un être vivant) mais conservent une relation de parenté avec leurs ancêtres vivants dont on peut suivre la trace à travers l'évolution de leur ADN.

110 Les virus pourraient être une forme de vie parasite née en même temps que la vie telle que nous l'entendons et ayant co-évolué avec elle mais sans lui être apparentée.

d'autres planètes ou créer en laboratoire n'a aucun lien de parenté avec celle que nous connaissons et ne peut donc pas être incluse sur le même arbre phylogénétique. Par conséquent, s'il fallait développer une classification universelle des êtres vivants, celle-ci ne pourrait pas s'appuyer sur un principe de continuité historique et d'évolution darwinienne à partir d'un ancêtre commun. La classification phylogénétique ne cesserait pas d'être valable, mais d'autres méthodes deviendraient nécessaires.

Que se passerait-il si la vie ou les êtres vivants n'étaient pas un genre naturel et si l'appartenance d'un objet à cette classe n'était pas inscrite dans le « livre de la nature », mais dépendait en grande partie d'un choix humain ? Comme pour les planètes et le statut de Pluton, la réponse est qu'il ne se passerait sans doute rien d'un point de vue scientifique, si ce n'est peut-être un changement de perspective pouvant favoriser la formulation de questions et d'hypothèses nouvelles.

Les conséquences seraient par contre plus importantes dans les domaines de la culture, de l'éducation, de la bioéthique, car le statut d'être vivant soulève des questions et il véhicule des représentations et des valeurs qui ne s'appliquent pas à la matière inanimée. La création artificielle de vie organique, de vie informatique, de robots autonomes et d'agents intelligents, ainsi que les questions concernant le début et la fin de la vie humaine sont des enjeux sociétaux de première importance. Par conséquent, s'il n'existe pas de frontières préétablies entre la vie et la non-vie, les scientifiques ne devraient peut-être pas être les seuls à prendre ce type de décisions. Comme dit Evelyn Fox Keller, c'est une décision humaine au sens large :

> Elle dépend de nos besoins locaux et de nos intérêts, de nos estimations des coûts et des bénéfices à procéder de la sorte et aussi, bien entendu, de notre plus large tradition culturelle et historique. La possibilité même d'inclure ces nouvelles entités aurait semblé absurde aux personnes vivant il y a peu ; et cela me semble toujours absurde. Mais cela ne veut pas dire pour autant que nous ne le ferons pas ou que nous ne devrions pas le faire. La question « qu'est-ce que la vie ? » est historique et traitable seulement dans les termes des catégories que, nous humains, voulons défendre, selon les différences que nous choisissons d'honorer, et non en des termes logiques, scientifiques ou techniques. C'est en ce sens qu'il faut voir la catégorie des vivants comme une sorte de famille — une famille qui est délimitée de manière humaine plutôt que "naturelle". (Keller, 2007, p. 50)

Par ailleurs, dire que les êtres vivants et les planètes ne sont pas une classe naturelle mais humaine ne nous pousse pas forcément à nier leur existence et leur réalité, ni celles de leur classe. Nous allons pas aborder en détail le problème du réalisme, mais un détour par la philosophie des artefacts peut nous aider à esquisser quelques idées à ce propos.

En effet, les artefacts et leurs catégories sont eux-aussi souvent conçus comme étant des produits des intentions et des croyances humaines, de sorte qu'ils ne sont pas indépendants de l'esprit. Malgré tout, certains sont réalistes à leur propos, c'est-à-dire qu'ils les reconnaissent comme une classe réelle et pas comme une classe nominale ou conventionnelle. L'une des stratégies consiste à dire que les classes d'artefacts ont une essence ou une structure interne commune, ou encore une histoire fonctionnelle commune (Elder, 2007). Cela revient à essayer de gommer la distinction entre les genres naturels et les classes artificielles (Soavi, 2009).

Carrara et Vermaas (2009) défendent eux-aussi l'idée que les artefacts ont des classes réelles dont l'essence est constituée (du moins en partie) par leur fonction, et ils montrent que cela rend possible des distinctions métaphysiques plus fines que celles des experts, c'est-à-dire que les principales conceptions des fonctions techniques n'arrivent pas à rendre compte du fait que deux artefacts en tout identiques mais qui n'ont pas la même fonction appartiennent à des classes réelles différentes. Or, cette conclusion est également valable pour les fonctions biologiques, c'est-à-dire que deux organes homologues constitués de tissus similaires et placés dans des structures anatomiques similaires (par exemple : les muscles de l'aile d'un petit oiseau visant sur une île orageuse et ceux d'un oiseau similaire vivant sur le continent) n'appartiennent pas à la même classe s'ils ont des fonctions différentes.

Une autre stratégie, adoptée par Amie Thomasson (2007, p. 72), consiste à dire que les classes artificielles sont réelles dans la mesure où elles sont créées et déterminées par des intentions humaines. La stratégie à adopter pour déterminer quelles entités devraient être admises dans notre ontologie, dit-elle, est la suivante : quelle que soit la classe d'entité, il faut d'abord déterminer quelles sont les conditions à remplir pour qu'il y ait des entités de ce type, et ensuite essayer de voir si les critères en question sont remplis. Or, pour qu'il y ait des artefacts et des classes d'artefacts, il faut qu'il y ait des gens avec l'intention de créer des objets d'une classe donnée, et ces intentions doivent remplir certains critères de succès :

> « According to the criteria built into the idea of something being an artifactual kind term, what must be the case for there to be artifacts and artifactual kinds? There must, as we have seen earlier, be people with certain intentions to create objects of a given kind, where these intentions are substantive and involve certain success criteria that control their activity, and they must be largely successful in executing their intentions. Do we have reason to think this is ever done? Barring radical conspiracy theories, of course we do. » (2007, p. 72)

Cette stratégie, continue l'auteur, ne permet pas d'inclure n'importe quoi dans notre ontologie, puisque des entités comme le phlogistique et les fantômes ne remplissent pas les conditions qui correspondent à leur classe. Pour qu'il y ait des fantômes, il faudrait qu'il y ait des personnes mortes qui reviennent avec une forme visible et spatiotemporelle, mais pas matérielle, et qui interagissent causalement avec le monde (en affectant les yeux de certains témoins, en déplaçant des objets, etc.). Or, nous avons de bonnes raisons de penser que de telles choses n'existent pas. Par contre, nous avons de bonnes raisons de penser que les conditions nécessaires pour qu'il y ait des chaises sont effectivement remplies.

Les êtres vivants ne sont vraisemblablement ni une classe naturelle ni une classe artificielle, et nous ne savons pas quelles sont les conditions à remplir pour que l'on puisse admettre l'existence d'êtres vivants. Par ailleurs, on peut inventer des catégories arbitraires dont les conditions d'existence sont remplies, comme celle des personnes dont le prénom contient un *a* et qui sont nées un lundi. Cependant, de telles catégories arbitraires n'ont sans doute aucune capacité explicative. La question est de savoir dans quelle mesure la classe des êtres vivants est arbitraire ou dictée par la nature, et dans quelle mesure elle est explicative.

Une stratégie différente, proposée par Diego Lawler et Jesús Vega (2011), consiste à renoncer aux « classes » en faveur de la notion de « ressemblances de famille » tirée de Wittgenstein. Ces auteurs rejettent l'essentialisme associé à la notion de classe, mais ne renoncent pas à une forme de réalisme à propos des artefacts. La constitution des artefacts ne repose pas sur une essence, ni une structure ou une histoire commune, ni ne peut être fondée de la même manière que les genres naturels, mais elle ne dépend pas non plus des intentions d'un créateur ni même du groupe social auquel il appartient ; elle est forgée et structurée par nos formes de vie :

> « Optar por dictaminar lo que un artefacto es según el vocabulario de las clases naturales significa forzar la posibilidad de una ciencia de lo artificial demasiado cercana y contaminada de los problemas de la ciencia de lo natural. Argumentar que los artefactos constituyen clases humanas y que éstas se entienden en base a los conceptos e intenciones de sus hacedores conlleva modelar el ámbito de lo artificial bajo la influencia de la idea de creación divina. A diferencia de estas opciones, sustituir el vocabulario de clases por la noción de «parecidos de familia» y recurrir a nuestras prácticas humanas de lidiar con nuestros artefactos para caracterizar lo que son, supone pensar en lo que ellos son a partir de los conceptos que aplicamos, los juicios apropiados que realizamos sobre ellos, las inferencias que trazamos, el cotejamiento de las genealogías que trazan sus linajes, etcétera. Esto puede parecer demasiado blando para decir qué son los artefactos, pero justamente implica el reverso de este parecer, puesto que nuestras prácticas sobre lo artificial son en sí mismas,

> como si dijéramos, el medio de la complejidad conceptual del ámbito de lo artificial. En esta estrategia no sólo los artefactos no pierden su rostro, sino que la ontología se acompasa a nuestras consideraciones normativas. » (Lawler & Encabo, 2011, p. 145)

Une autre manière d'aborder la question consiste à se demander, non pas si les entités et leurs classes sont indépendantes de l'esprit, ou si elles existent réellement dans le monde, mais si elles sont naturelles — c'est-à-dire qu'elles correspondent à la manière dont le monde est structuré — et si elles sont explicatives — c'est-à-dire si elles jouent un rôle théorique :

> « Philosophers have been concerned not only with whether the posits of science exist mind-independently but also with whether these posits are "appropriately special" rather than somewhat arbitrary. In particular there has been a concern about whether our scientific posits "carve nature at its joints," about whether there is something in the nature of the world that, in some sense, determines our categorization of it. I take this to be a concern about whether the kind of entity posited by a theory plays a causally significant role, whether it is partly because an entity is of that kind that it has the characteristics and behavior that it has. Theories need to posit such so-called "natural" kinds if the theories are to be genuinely explanatory. And Realists are likely to take it for granted that their paradigm entities — for example, cats and planets — are indeed of natural kinds. » (Devitt, 2011, p. 158)

De nouveau, la question ici est de savoir si les catégories « planète » et « vivant » sont explicatives et si elles sont déterminées par la nature ou par les intérêts humains.

3. Définitions fonctionnelles du vivant

De nombreuses définitions de la vie sont liées directement ou indirectement à la notion de fonction. Aristote, par exemple, considérait non seulement que la vie consiste à se nourrir par soi-même, à se développer et à périr (*De l'âme*, II, 1, 412a), lesquelles sont des propriétés fonctionnelles, mais il plaçait la fonction elle–même, liée à la forme et à l'âme, au cœur de sa biologie (Pichot, 1993). Plus près de nous, Jacques Monod (1970) affirmait que les propriétés les plus générales du vivant sont la téléonomie, la morphogenèse autonome et l'invariance reproductive. John Maynard Smith (1986) proposait quant à lui deux critères pour caractériser les êtres vivants : le métabolisme et la possession de fonctions. Il est l'un des seuls, avec Mark Bedau (2007) et Carl Sagan (1970), à mentionner explicitement les fonctions dans sa définition de la vie.

3.1. La formulation de Sagan

La formulation de Sagan est particulièrement intéressante dans la mesure où elle reprend la plupart des activités fonctionnelles citées par les autres définitions :

> [Life is] the state of a material complex or individual characterized by the capacity to perform certain functional activities, including metabolism, growth, reproduction, and some form of responsiveness and adaptation. Life is further characterized by the presence of complex transformations of organic molecules and by the organization of such molecules into the successively larger units of protoplasm, cells, organs, and organisms.

À la suite de cet auteur, on peut classer les différentes définitions en six grandes catégories : physiologique, métabolique, biochimique, thermodynamique, darwinienne et autopoïétique.

(a) Définition physiologique

Les êtres vivants sont des systèmes capables de réaliser un certain nombre de fonctions (alimentation, métabolisation, excrétion, respiration, croissance, mouvement, reproduction, etc.). Mais cette définition ne permet pas de distinguer les êtres vivants des artefacts. Pour la rendre praticable, il faudrait établir une liste de fonctions qui soient à la fois communes à tous les êtres vivants et propres à eux seuls. Et encore, cette liste ne serait sans doute pas suffisante, car si nous étions capables de fabriquer un robot mécanique remplissant toutes ces fonctions, serait-il vivant ?

(b) Définition métabolique

Les êtres vivants sont des objets aux contours définis qui échangent continuellement du matériel et de l'énergie avec leur environnement sans perdre leurs propriétés. Mais une flamme pourrait satisfaire cette définition[111]. Par ailleurs, des graines, des spores et même des animaux

111 Le fait d'avoir des limites spatiales bien définies n'implique pas nécessairement la possession d'une membrane. Cette dernière est un bon moyen d'éviter la dilution des réactifs, mais ce n'est pas le seul. Aux origines de la vie sur Terre, des pores et des interstices rocheux ont peut-être fait office de membrane minérale pour les premières biomolécules. D'autres modes de confinement sont envisageables. Par exemple, une goutte d'huile dans l'eau a une frontière précise sans avoir de membrane. Le fait qu'une flamme n'ait pas de membrane ne constitue donc pas une objection valable. Le feu répond par

(les tardigrades) peuvent rester inertes (sans activité métabolique connue) durant de longues périodes et se « réveiller » lorsque les conditions sont propices pour retourner à une activité normale[112]. Une spore bactérienne enfermée dans un cristal de sel a ainsi été « ramenée à la vie » après 250 millions d'années d'inactivité (Vreeland, Rosenzweig, & Powers, 2000).

(c) Définition biochimique

Les êtres vivants sont des systèmes basés sur la chimie du carbone qui contiennent une information héréditaire encodée dans des molécules d'acide nucléique et qui sont dotés d'un métabolisme contrôlé notamment par des enzymes. Mais des créatures extraterrestres pourraient avoir des systèmes biochimiques différents basés par exemple sur le silicium ou l'arsenic (National Research Council, 2007; Wolfe-Simon et al., 2010), voire être dépourvus d'information héréditaire encodée (ou alors encodée différemment), tout en ayant à peu près les mêmes propriétés fonctionnelles que les organismes terrestres.

(d) Définition thermodynamique

Les êtres vivants sont des systèmes ouverts (qui échangent de la matière et de l'énergie avec leur environnement) qui se situent hors de l'équilibre thermodynamique et dont l'entropie locale reste stable ou décroît[113]. Mais des phénomènes similaires ont lieu dans l'espace ou sur d'autres corps célestes, en l'absence de vie, avec la production de molécules de plus en plus complexes — y compris des sucres et des acides aminés — lesquelles ont justement rendu possible l'apparition des premières formes de vie sur Terre.

---

ailleurs aux principaux critères du vivant : il naît, croît, meurt, se propage, se reproduit (par le biais d'étincelles) et répond à des stimuli externes comme le vent. Il est doté d'un métabolisme consumant de l'oxygène et produisant des déchets.

112 Si l'on ne veut pas parler de mort et de résurrection ni de vie à l'état latent, on peut toujours défendre l'idée que les individus « avant » et « après » ne sont pas le même, le second étant « né » après la « mort » du premier et à partir de ses restes.

113 Autrement dit, quoi que les deux formulations ne soient pas strictement équivalentes : dont le degré d'ordre reste stable ou augmente.

(e) Définition génétique ou darwinienne

La vie est un système capable d'évolution par sélection naturelle[114]. C'est une définition fonctionnelle qui a l'avantage d'être suffisamment générale pour s'appliquer à des formes de vie différentes de celles que nous connaissons, mais qui a l'inconvénient de s'appliquer aussi à des objets que nous ne reconnaissons pas comme vivants : les cristaux d'argile, les virus biologiques, les virus informatiques, etc.

(f) Définition autopoïétique

Les êtres vivants sont des systèmes autonomes individuels, autoproduits et automaintenus qui conservent constantes leur organisation et leur identité. Cette définition ne fait aucune référence à la reproduction ni à l'évolution et elle est suffisamment générale pour s'appliquer aussi bien à des bactéries qu'à des écosystèmes.

Chacune de ces définitions, s'applique à un domaine d'objets différent : ce qui est vivant pour l'une ne l'est pas forcément pour l'autre ; et même en limitant leur portée à la vie sur Terre, aucune ne parvient à saisir exactement ce qui en fait la spécificité. Pourtant, les êtres vivants que nous connaissons manifestent une très grande unité dans leur diversité, laquelle s'explique par leur ascendance commune. La difficulté ici ne consiste pas à saisir ce que tous ont en commun (chimie du carbone, ARN/ADN, sélection naturelle, etc.), mais à situer quelque part la frontière qui les sépare d'autres objets partageant les mêmes caractéristiques. Et la difficulté s'accroît lorsque l'on cherche à donner une définition qui ne se limite pas aux formes de vie connues, mais qui embrasse également celles que nous pourrions découvrir ici ou ailleurs.

3.2. La formulation de Joyce

La recherche de vie extraterrestre implique de savoir distinguer le vivant du non-vivant n'importe où dans l'univers et indépendamment des formes matérielles que nous connaissons sur Terre. Une définition générale doit donc s'abstraire des détails physico-chimiques susceptibles de varier d'un endroit à l'autre.

Les définitions fonctionnelles présentent de ce point de vue l'avantage de la multi-réalisabilité : les propriétés fonctionnelles peuvent être satisfaites par de multiples systèmes physicochimiques très différents les uns des autres. Par exemple, l'une des définitions les plus populaires dans le milieu de la recherche sur les origines de la vie et sur la vie extraterrestre est une définition fonctionnelle :

114 On peut élargir la définition de façon à y inclure des mécanismes évolutifs non proprement sélectifs comme la dérive génétique.

> La vie est un système chimique auto-maintenu capable d'évolution darwinienne[115]. (Joyce, 1994).

En précisant qu'il s'agit d'un processus chimique, la définition de Joyce (dite de la NASA) exclut toute forme de vie *in-silico* (qui autrement pourrait la satisfaire, car la définition ne mentionne pas directement le métabolisme). En indiquant que le système est auto-maintenu (*self-sustained*), elle fait référence au fait que les êtres vivants contiennent tous les moyens nécessaires pour mener à bien leur constante auto-production, ce qui revient à exclure les virus. Et à travers la notion d'évolution darwinienne elle subsume les processus d'auto-reproduction, de continuité à travers une lignée historique, de variation génétique et de sélection naturelle. Cependant, prise au pied de la lettre, la définition n'est pas applicable aux êtres vivants individuels — car ceux-ci ne sont pas capables d'évolution darwinienne —, ni aux populations ou aux espèces qui ne sont pas, à proprement parler, des systèmes chimiques auto-maintenus et dont la capacité d'évolution aboutit à la spéciation, c'est-à-dire à la disparition de l'entité « mère » au profit de deux entités « filles ». Comme le rappelle Aaron Goldman (2011), de la NASA, l'interprétation que font les astrobiologistes de cette définition, laquelle est issue de leur discipline, les amène à dire que nous ne connaissons à l'heure actuelle qu'une seule forme de vie : la Biosphère (hypothèse Gaïa). Dans ce contexte, ceux que nous appelons « êtres vivants » ne sont en vie que dans la mesure où ils participent de cette entité supérieure. Et, dans ce contexte, la discussion sur le fait que les virus soient ou pas des êtres vivants est dénuée d'intérêt : les virus ne sont pas la vie, mais nous non plus (A. Goldman, 2011).

4. La définition de la vie peut-elle être arbitraire ?

Si la définition du vivant doit être fonctionnelle, le sens que l'on prête au concept de fonction n'est pas indifférent. Une interprétation à la Cummins peut signifier que les individus incapables de se reproduire ne possèdent pas la fonction correspondante et ne remplissent pas les conditions de la définition, ce qui impliquerait qu'ils ne sont pas vivants. Une interprétation étiologique, au contraire, dirait que ces individus sont des

115 « Life is a self-contained chemical system capable of undergoing Darwinian evolution. » D'après Luisi (2006, p. 47), il ne s'agissait à l'origine que d'une définition de travail, une perspective opérationnelle dans le cadre du programme d'exobiologie de la NASA.

êtres vivants dont la reproduction est dysfonctionnelle. Dans le premier cas, les êtres vivants sont définis par leurs capacités actuelles. Dans le second, ils sont définis par leur histoire reproductive ou, plus généralement, par leur appartenance à un type. Alors que la définition du concept de fonction dépend de celle du vivant, cette dernière dépend à son tour de l'interprétation du concept de fonction.

La définition de Joyce combine plusieurs éléments présents dans les définitions antérieures. Elle mélange notamment une condition systémique avec une condition darwinienne. C'est sans doute une bonne combinaison, mais ce n'est pas la seule envisageable. On pourrait par exemple affirmer que la vie est un système ouvert hors d'équilibre capable d'auto-réplication. On pourrait aussi définir la vie en termes de fonctionnalités telles que le traitement autonome de l'information et la sensibilité à l'environnement, abstraction faite de l'implémentation matérielle et des capacités reproductives des entités en question. Trifonov (2011) a analysé 123 définitions pour en extraire le plus petit commun dénominateur qui se réduit à cette définition minimale : « *life is self-reproduction with variations* », mais la vingtaine de spécialistes invités à répondre à son article ont exprimé leur désaccord pour des raisons diverses (Trifonov, 2012).

Pourquoi les définitions de Sagan, de Joyce ou de Trifonov seraient-elles meilleures que d'autres ? Quels critères d'évaluation emploie-t-on pour juger de la validité et de la valeur d'une définition du vivant ? Et les différentes définitions qui ont été proposées jusqu'à maintenant, yvant compris celle de Joyce, sont-elles autre chose que des collections plus ou moins arbitraires de traits communs aux entités que nous reconnaissons sur Terre comme des êtres vivants ?[116] D'après l'un des textes classiques de l'exobiologie, le seul consensus auquel on soit parvenu en biologie théorique, dont l'un des objectifs principaux est la définition de la vie, est qu'une telle définition doit être arbitraire (Lederberg, 1960, p. 394).

La définition des planètes est-elle arbitraire ? Oui et non. Si l'on accepte l'idée qu'une définition vise avant tout à capturer un concept, alors il se trouve que tout le monde ne partage pas forcément le même concept. Nous avons vu que pour certains astronomes le terme « planète » ne fait pas référence à un concept scientifique, mais à une

116 À vrai dire, on peut se demander si les formulations précédentes sont des définitions — à proprement parler — ou seulement des listes de critères. Les critères sont en quelque sorte des règles servant à identifier et à différencier de manière systématique des faits ou des entités. Mais définir ce n'est pas seulement fixer les limites d'une chose ou d'un concept, c'est aussi en extraire l'essence ; c'est déterminer ce que la chose ne peut pas ne pas être. Selon Bedau (2007, p. 8), l'essence de la vie est la faculté à développer une évolution sans limite. Relativement à cette essence, les trois fonctions critiques qui d'après lui caractérisent tous les êtres vivants, à savoir un métabolisme, des gènes et un conteneur, doivent être conçues comme des critères de la définition.

notion culturelle, et que les autres ne sont pas non plus insensibles à des considérations extrascientifiques (sentimentales, économiques et autres). Sur le plan scientifique, les astronomes sont partagés entre trois perspectives, celle des caractéristiques intrinsèques (géophysiques), celle du contexte (circonstances orbitales) et celle de l'histoire ou des origines (cosmogonie), auxquelles ils assignent des pondérations différentes qui reflètent en partie les intérêts scientifiques de leurs disciplines respectives. La définition de l'UAI est donc arbitraire au sens où elle résulte du choix d'un concept parmi d'autres au terme d'une négotiation tendue entre de nombreux spécialistes en désaccord sur des questions essentielles. Nonobstant, on peut dire aussi que la définition finalement retenue n'est pas arbitraire dans la mesure où elle capture un concept rationnellement justifié et scientifiquement pertinent qui reflète certaines structures du monde naturel.

La définition de l'UAI est stipulative et conventionnelle. Elle ne répond pas à la question « Qu'est ce qu'une planète ? » mais à « Quels objets allons-nous qualifier de "planètes" ? ». La première question est essentialiste, la seconde est pragmatique. L'une n'exclut pas l'autre, car à défaut de savoir ce qu'est un objet, on peut toujours se mettre d'accord pour l'identifier et le caractériser aussi précisément que possible, quoique de façon provisoire, à des fins de communication et de recherche. Une question sous-jacente reste posée : Les planètes existent-elles indépendamment des connaissances que nous en avons ? Sont-elles un type d'objet naturellement distinct que la science nous amenène à découvrir, ou ne sont-elles qu'un mode de catégorisation humaine produit de la culture et de la science ?

La biologie, en tant que discipline scientifique autonome, semble être née de l'intuition qu'il existe une unité sous-jacente à tous les êtres vivants, une essence commune que la science nous amènerait à découvrir. Son émergence « *implique la conviction que l'ordre vivant a une originalité phénoménologique, voire ontologique, requérant une démarche de connaissance spécifique* » (Fagot-Largeault, 2002, p. 495). L'étude des formes de vie terrestre a montré qu'elles partagent de nombreuses caractéristiques explicables par leur origine commune et par leur histoire évolutive. Depuis cette perspective, le but d'une définition universelle de la vie est d'identifier, parmi les propriétés des organismes connus, celles qui sont partagées par tous les êtres vivants possibles. Chaque tentative de définition repose ainsi sur un petit nombre de propriétés qu'elle considère essentielles et qui délimitent un ensemble d'objets possibles dont les organismes terrestres font partie (Fig. 12). S'il existe une distinction naturelle entre les êtres vivants et les objets inertes, et si cette distinction repose sur un ensemble de propriétés communes aux êtres vivants en général, alors notre conception de la vie doit rendre compte de ces propriétés, et la définition correspondante doit capturer au mieux cette conception. Autrement dit, le domaine d'objets définis par elle doit s'ajuster au mieux

à l'ensemble naturel des êtres vivants possibles.

Mais nous avons vu qu'il est possible de découvrir ou de créer, pour chacune des conceptions concurrentes, des objets qui lui soient conformes. Dans l'exemple de la Fig. 12, cela voudrait dire qu'il est possible de trouver ou de fabriquer des objets dotés respectivement de l'une ou l'autre des propriétés *V*, *I* et *E*, ainsi que de leurs combinaisons. Si tel est le cas, alors chacun de ces objets pourrait être considéré comme vivant selon l'une des conceptions, mais pas selon d'autres. Tous ces ensembles d'objets pourraient par ailleurs se prêter à des généralisations inductives. Par conséquent, à moins de considérer qu'ils sont tous également vivants, ce qui impliquerait l'absence de propriétés communes à la vie en général, comment justifier le choix d'une des options concurrentes ? À défaut de savoir identifier — objectivement — les objets qui appartiennent à l'ensemble naturel des êtres vivants (dont on présuppose jusqu'ici l'existence), le choix d'une conception plutôt qu'une autre apparaît pour le moins délicat.

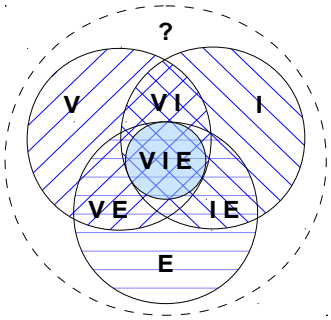

Figure 12: Domaines d'objectivité de plusieurs définitions hypothétiques de la vie. La vie terrestre, en bleu au centre, se caractérise par trois propriétés : V, I et E. Chacune des définitions candidates s'appuie sur une ou sur deux de ces propriétés.

De plus, la possession de propriétés communes n'est pas suffisante pour constituer un genre naturel. Les objets rouges de masse 1 kg, par exemple, ne sont pas un genre naturel. Si les êtres vivants étaient un genre naturel, alors une définition listant leurs propriétés communes ne saurait donc suffire à les caractériser. Cette définition devrait s'appuyer en outre sur une théorie de la vie expliquant les raisons d'être de l'ensemble singulier de propriétés sur lequel elle repose. De même que les propriétés de l'eau s'expliquent par sa composition chimique, les propriétés de la vie pourraient s'expliquer soit par une composition, une structure ou un processus sous-jacents, soit par une histoire causale particulière ayant conduit à les regrouper. Dans ce dernier cas, les êtres vivants — du moins sur Terre — seraient une « grappe naturelle » (*cluster kind*) caractérisée par une ou plusieurs familles de propriétés groupées de façon contingente par la nature (Boyd, 1999; Millikan, 1999). Ces familles de propriétés seraient groupées ensemble à travers le temps à cause de la présence de certaines propriétés favorisant la présence des autres, ou à cause de mécanismes internes ou externes tendant à faire coïncider ces propriétés. Dans cette perspective, les êtres vivants pris dans leur universalité pourraient ne pas former un seul genre naturel mais plusieurs, chacun étant lié par une

histoire particulière. C'est-à-dire que les formes de vie terrestre et celles qui existent potentiellement sur d'autres planètes pourraient constituer des genres naturels — ou plutôt des grappes naturelles — distincts. Par ailleurs, étant donné que la notion de genre naturel est sujette à discussions, il est préférable de parler de distinction naturelle plutôt que de genre naturel, car les êtres vivants pourraient ne pas former un genre naturel tout en étant naturellement distincts des objets inanimés.

Le problème est que nous n'avons aucune théorie de la vie ou nous en avons trop, à savoir la théorie autopoïétique (Damiano & Luisi, 2010; Varela et al., 1974) et la théorie darwinienne (Pross, 2011), pour n'en citer que deux. Celles-ci permettent d'expliquer des phénomènes naturels caractéristiques de ce que nous appelons la vie et d'établir entre les objets naturels des distinctions scientifiquement pertinentes, mais comme pour les planètes il y a plusieurs bonnes options disponibles.

La biologie doit-elle répondre à la question « Qu'est-ce que la vie ? » ou peut-elle se contenter de choisir « Quels objets qualifier de "vivants" ? ». Qu'il y ait une essence de la vie, ou que la vie soit une propriété commune et spécifique aux êtres vivants pris dans leur universalité, cela reste un présupposé. Celui-ci n'est pas nécessaire et nous avons de bonnes raisons de ne pas y croire. Or, si l'appartenance d'un objet au domaine du vivant n'a pas de réponse préétablie, et si les êtres vivants se caractérisent par la fonctionnalité de leurs parties, alors le fait qu'une structure ou un processus aient ou pas une fonction n'a pas non plus de réponse préétablie.

5. La transition de l'inerte au vivant a-t-elle lieu dans la nature ou dans le regard ?

5.1. Du langage de la physico-chimie à celui de la biologie

De nombreux chercheurs, comme Harold Morowitz (1993, 1999; Morowitz & Smith, 2007), pensent que l'apparition de la vie sur Terre était sans doute inévitable et qu'elle doit par conséquent apparaître sur toute autre planète présentant des caractéristiques similaires. L'approche constructive en biologie synthétique consiste ainsi à partir de la physico-chimie pour essayer de remonter par complexification croissante jusqu'au domaine biologique afin de comprendre la nature des processus impliqués et en dégager éventuellement des lois universelles. Cette approche correspond à une conception systémique du vivant dont nous avons vu que le principal défaut est ce que Davies (2001) appelait, à propos des fonctions, la « promiscuité » de la théorie. C'est-à-dire qu'elle attribue avec trop de libéralité de la vie et des fonctions à des objets qui intuitive-

ment n'en ont pas, car elle est incapable de situer la frontière séparant un domaine d'un autre. En effet, quand on essaie de suivre le cours de la complexité en allant des processus physico-chimiques simples vers les objets biologiques, le problème est de savoir où se situe la frontière (si frontière il y a) et à quel niveau s'effectue la transition. Il ne fait aucun doute que si la vie était une propriété émergente de l'organisation de la matière, la construction en laboratoire d'une cellule vivante semi-synthétique en serait la démonstration évidente. Mais cela n'implique pas que nous sachions identifier le moment à partir duquel un système commence à être vivant.

Les systèmes physiques relèvent de la physique et il faut les étudier depuis la perspective qui leur correspond, avec les concepts et les méthodes de cette discipline. À mesure que les systèmes deviennent plus complexes, les outils d'analyse sont aussi plus puissants et plus sophistiqués, tout en restant dans le cadre de la physique, car cette science a une « compétence universelle » : des particules élémentaires aux grandes structures de l'univers, rien de ce qui est dans la nature n'est censé lui échapper. Un système physique reste un système physique quelles que soient ses propriétés émergentes et son degré de complexité. Or, la vie et les fonctions sont des concepts étrangers à la physique. Une description complète des processus physico-chimiques ayant conduit à la formation des premiers organismes à partir de molécules simples ne mentionnerait à aucun moment l'apparition de la vie et des fonctions, car ces notions n'ont aucune signification physique. Par conséquent, d'un point de vue strictement physique, la transition de l'inerte au vivant n'existe pas.

En revanche, il y a bel et bien une transition au niveau du langage et des méthodes employés pour les décrire. Bien que la physique soit en principe compétente pour étudier des systèmes aussi complexes que les bactéries, c'est généralement la biologie qui s'en charge. Et bien que les virus ne soient pas considérés comme vivants, ils sont aussi de son ressort. Il est difficile de dire où se situent les frontières de chacune des deux disciplines, d'autant plus qu'il existe des spécialités se situant à l'interface, comme la biochimie, la biophysique, la physique de la « matière molle », etc. Mais ce qui est sûr, c'est que l'attribution de vie et de fonctions à certains objets marque une rupture entre le modèle épistémologique de la physique et celui de la biologie. Et les difficultés rencontrées pour identifier les frontières du vivant sont peut-être liées au fait que l'on recherche une hypothétique transition entre deux états de la matière là où il n'existe en réalité qu'un changement de discours.

Les vésicules autopoïétiques de Luisi et Varela sont peut-être vivantes, peut-être pas ; de même que les virus. Cela fait-il une différence ? Apparemment non puisque ces mêmes objets peuvent être étudiés aussi bien du point de vue de la physique que de la biologie et aucune des deux disciplines n'a besoin de se préoccuper de leur statut vital. Or, si la vie était une propriété émergente de la matière, les objets en question

posséderaient ou ne posséderaient pas ladite propriété. Et à moins qu'elle ne fût incolore, inodore et insipide, cela devrait faire une différence. Au minimum, cela devrait permettre de développer un test objectif permettant de dire si un objet possède ou pas la propriété en question. Jusqu'à présent, aucun des critères proposés n'a réussi à s'imposer ; non pas parce qu'ils ne fussent pas bons, mais parce qu'aucun n'a montré être meilleur que les autres. Peut-être parce que la propriété en question (si elle existe) n'est pas de celles dont on peut faire un détecteur embarqué dans une sonde spatiale ou implémenté sur une biopuce.

Le basculement du discours de la biologie à celui de la physique est peut-être moins lié à un problème d'émergence (apparition d'un nouvel état de la matière) que de pertinence (nécessité d'un nouveau mode d'explication). C'est-à-dire qu'à partir d'un certain degré de complication[117], les outils traditionnels de la physique cesseraient d'être adaptés, et ceux de la biologie prendraient le relais. Ou plutôt : la pertinence (l'utilité, l'efficacité, la practicalité) des outils de la physique serait de plus en plus limitée tandis que celle des outils de la biologie augmenterait. Il en va de même pour d'autres branches des sciences de la nature. On peut expliquer le comportement et l'évolution de n'importe quel système physique à partir de la mécanique quantique, mais à mesure que le nombre de particules du système augmente il devient de plus en plus difficile de le suivre et l'usage d'outils alternatifs comme la mécanique statistique devient de plus en plus nécessaire. Il n'y a pas un nombre précis de particules qui forcent le passage de l'un à l'autre, pas plus qu'il n'y a un nombre absolu de gènes caractérisant une cellule vivante minimale[118]. Car la différence n'est pas seulement dans les objets mais aussi dans les pratiques : c'est une question d'efficacité.

5.2. Connaissance et aspectualité

Si la distinction entre la vie et la non-vie n'est pas dans la nature mais dans nos représentations de la nature, c'est-à-dire dans notre regard, alors on peut légitimement penser qu'elle est arbitraire. Et si elle repose

117 Nous avons vu que la notion de complexité n'était pas clairement définie et que l'identification d'un seuil de minimalité est assez délicate ; par conséquent, le degré de complexité d'un système à partir duquel les outils de la physique commencent à être moins pertinents que ceux de la biologie doit plutôt être compris au sens d'un degré (plus ou moins subjectif) de complication.

118 Cela n'est pas toujours vrai. En astronomie, les équations de Newton permettent de trouver une solution analytique au mouvement de deux corps en interaction gravitationnelle, mais à partir de trois corps il devient nécessaire d'opter pour une résolution approchée à travers la théorie des perturbations ou l'analyse numérique. Il y a donc bien une frontière (le passage de deux à trois corps) forçant l'utilisation d'un outil différent.

principalement sur l'intuition et la tradition historique, alors on peut légitimement penser qu'elle est subjective. De fait, les limites du vivant et de l'inerte varient en fonction de l'époque, de la culture et de l'âge des personnes interrogées. La biologie — science de la vie — y perdrait alors l'objet sur lequel elle a été fondée et qui fait son unité. Elle deviendrait la science des objets physico-chimiques qui partagent un lien de parenté avec les *homo sapiens* en vertu d'une origine commune et de l'évolution darwinienne[119].

Seulement voilà, ce point de vue ne permet pas de comprendre pourquoi et comment la distinction prétendument arbitraire entre les êtres vivants et les objets inanimés a pu, jusqu'à aujourd'hui, s'avérer aussi féconde et aussi pertinente scientifiquement. La classification phylogénétique du vivant est une manière parmi d'autres de le classer, mais elle est incontestablement meilleure que celle de l'encyclopédie chinoise qu'imagine Borges[120]. Meilleure parce qu'elle repose sur un principe rationnel qui la rend intelligible, et aussi parce qu'elle contribue à la connaissance des objets ainsi classés, notamment sous la forme de généralisations inductives. Nous avons vu par ailleurs que la découverte ou la fabrication

---

119 Ou la science des objets qui partagent un certain nombre de mécanismes moléculaires centrés sur l'expression et la réplication de polymères informationnels. Ou encore la science des objets qui partagent les caractéristiques et les capacités spécifiques listées par Ernst Mayr (1997, p. 35).

120 La classification est celle que cite Foucault dans la préface de *Les mots et les choses* (1966). Elle appartient à l'essai « El idioma analítico de John Wilkins » (Borges, 1952). D'après George Lakoff (1987, p. 92) cette classification n'est pas si saugrenue qu'il y paraît, car on en trouve de similaires chez certains peuples non-occidentaux. Il rapporte à ce propos celle de la langue Dyirbal des aborigènes d'Australie : « *Whenever a Dyirbal speaker uses a noun in a sentence, the noun must be preceded by a variant of one of four words: bayi, balan, balam, bala. These words classify all objects in the Dyirbal universe, and to speak Dyirbal correctly one must use the right classifier before each noun. Here is a brief version of the Dyirbal classifcation of objects in the universe, as described by R.M.W. Dixon (1982): Bayi: men, kangaroos, possums, bats, most snakes, most fishes, some birds, most insects, the moon, storms, rainbows, boomerangs, some spears, etc. Balan: women, anything connected with water or fire, bandicoots, dogs, platypus, echidna, some snakes, some fishes, most birds, fireflies, scorpions, crickets, the stars, shields, some spears, some trees, etc. Balam: all edible fruit and the plants that bear them, tubers, ferns, honey, cigarettes, wine, cake. Bala: parts of the body, meat, bees, wind, yamsticks, some spears, most trees, grass, mud, stones, noises, language, etc. It is a list that any Borges fan would take delight in.* » Pour étrange que cette classification nous paraisse, on en trouve de comparables dans les langues occidentales : en français par exemple toutes les choses vivantes et inertes sont désignées par des noms sexués, et si l'on faisait la liste des choses masculines d'un côté et féminines de l'autre, l'arbitraire de la classification résultante n'aurait rien à envier à celle des aborigènes d'Australie.

de créatures vivantes sans ancêtre commun rendrait nécessaire l'élaboration de classifications non phylogénétiques. Il faudrait que ces dernières soient véritablement universelles et qu'elles permettent, elles aussi, des généralisations inductives répondant à des intérêts épistémiques particuliers. Donc, même si les différentes classifications du vivant ainsi que la distinction entre la vie et la non-vie n'étaient que des modes de représentation de la réalité, elles n'en seraient pas moins des sources potentielles de connaissance de cette réalité.

N'importe quel objet peut être décrit en termes physico-chimiques ou en termes biologiques. Si la distinction entre le vivant et l'inerte n'existe pas dans la nature mais dans le regard, alors la question n'est pas tant de savoir si ces descriptions sont correctes et si l'attribution de vie et de fonctions est vraie, mais plutôt si elles sont pertinentes scientifiquement. Et elles sont pertinentes si elles contribuent à notre connaissance de l'objet. Par exemple, essayer de décrire un caillou avec le langage et les méthodes de la biologie est vraisemblablement une perte de temps, alors que l'analyse des contraintes physiques qui s'exercent sur une membrane cellulaire peut aider à comprendre les mécanismes non génétiques de certaines pathologies comme la malaria (Abkarian, Massiera, Berry, Roques, & Braun-Breton, 2011). Le regard que nous portons sur un objet peut être conçu comme une stratégie interprétative dont la pertinence et la valeur se mesurent à l'aune de son succès.

La physique du siècle prochain sera peut-être capable de prédire le comportement d'un protozoaire sur la base de ses caractéristiques matérielles, mais elle ne pourra pas expliquer pourquoi certains de ses traits se rencontrent également chez d'autres organismes, des cyanobactéries aux homo sapiens. Et il est vraisemblable qu'aucun développement ultérieur des outils de la physique ne viendra changer cette situation. Les traits biologiques sont le fruit d'une contingence historique dont aucune loi physique ne peut rendre compte tout simplement parce que l'histoire n'est pas du ressort de la physicochimie[121]. Comme disait Jean Gayon, l'unité des phénomènes vitaux apparaît aujourd'hui sous deux visions antagonistes de ce champ de savoir, celle de la biologie moléculaire et celle de la théorie de l'évolution :

> Si l'on considère que, du point de vue théorique, l'essentiel est dans la physico-chimie de la vie, les sciences biologiques sont en voie d'être absorbées par les sciences générales de la matière, dont elles constituent en quelque sorte des provinces ou secteurs d'application. Toutefois l'on peut estimer, à l'instar de bon nombre des fondateurs de la biologie moléculaire, que cette discipline ne fournit pas à pro-

121 Le caractère aléatoire du processus évolutif n'est cependant pas incompatible avec l'existence de contraintes physico-chimiques qui le limitent et qui permettent d'expliquer certains cas de convergence évolutive comme la photosynthèse en C4 qui est apparue indépendamment plus de 60 fois.

> prement parler une « théorie » de la vie, mais un puissant outil technologique de description, qui sans doute apporte des indices supplémentaires et spectaculaires de l'unité des êtres et des phénomènes de la vie (comme l'avait déjà fait dans le passé la théorie cellulaire), mais n'épuise pas les questions de la diversité et du changement des formes vivantes. Dans cette perspective, c'est à la théorie de l'évolution organique, théorie de nature fondamentalement historique, que revient la responsabilité de comprendre l'unité propre du monde vivant. (Gayon, 1993, p. 47)

À chaque discipline correspond un type de question qui lui est propre et on ne peut pas reprocher à l'une de ne pas savoir répondre aux questions de l'autre. Biologistes, biochimistes, physiciens, astronomes, philosophes et informaticiens semblent se poser les mêmes questions : Qu'est-ce que la vie ? Où sont ses limites ? Qu'est-ce qui fait son unité ? etc. Pourtant, bien qu'elles soient formulées dans les mêmes termes, les questions ne sont pas tout à fait les mêmes et leurs réponses non plus. Il ne faudrait donc pas concevoir les différents types de définition (physiologique, thermodynamique, darwinienne, etc.) comme des alternatives concurrentes. Ce sont plutôt des réponses complémentaires à des questions complémentaires portant sur un objet qui est à la fois le même et un autre.

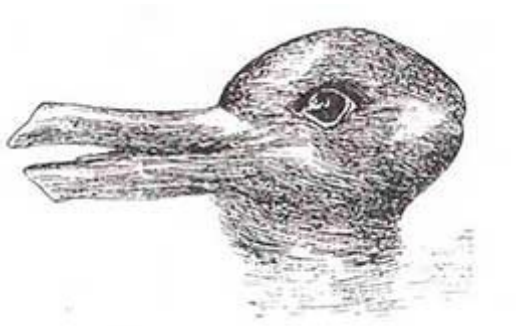

Figure 13: Image ambigüe que l'on peut voir comme un canard ou comme un lapin, mais jamais comme les deux à la fois.

Pour comprendre le caractère paradoxal des objets que nous appelons « êtres vivants » il faut penser aux images ambiguës comme la figure de Jastrow que l'on peut voir alternativement comme un canard ou un lapin et jamais comme les deux à la fois (Fig. 13). Alors que l'image est incontestablement la même, sa description ne l'est pas, et il n'y a pas une unique réponse correcte à la question « Qu'est-ce que c'est ? » La réponse dépend du point de vue que l'on adopte et il n'y a pas non plus un unique point de vue qui soit le bon. Cela étant, toutes les réponses ne se valent pas et certaines (canard, lapin) sont meilleures que d'autres

(éléphant, motocyclette). De la même manière, la question « Qu'est-ce que la vie ? » admet plusieurs réponses relatives chacune au point de vue adopté (physique, biochimique, évolutionniste, etc.) où certaines sont meilleures que d'autres et aucune n'est parfaite. De plus, la distinction vivant/inerte ressemble beaucoup à la dualité canard/lapin, car ce sont deux aspects complémentaires et simultanément exclusifs d'une même réalité. Dans un cas l'exclusion est perceptive, dans l'autre elle est méthodologique. Si l'on voit une entité comme un être vivant, on peut lui attribuer des fonctions ; si on la voit comme un système physico-chimique, non.

Par ailleurs, toutes les images ne sont pas ambiguës et tous les objets ne se prêtent pas à une double description physique-biologique. Ou plutôt : il n'est pas toujours aussi facile d'y voir plusieurs aspects également saillants. Il faudrait beaucoup d'imagination pour voir une chaise comme un être vivant ou pour y voir autre chose qu'une chaise. Car l'aspectualité n'est pas totalement arbitraire ni subjective, elle dépend de la configuration de la chose perçue au moins autant que de la personne et elle présente des degrés mesurables d'intersubjectivité. De fait, il n'est pas facile de fabriquer une image ambiguë, c'est-à-dire une image qui suscite deux interprétations possibles entre lesquelles le cerveau est incapable de choisir, et les exemples dans la nature sont rares. De même, les exemples d'objets naturels pour lesquels nous ne saurions dire s'ils sont vivants ou inertes ne sont pas monnaie courante. Pour tous les autres, la distinction est soit indifférente soit immédiate, car il en va de notre survie. Et de même que je ne choisis pas de voir un canard ou un lapin dans la figure de Jastrow, car autrement je pourrais choisir d'y voir un éléphant ou une motocyclette, je ne choisis pas non plus de voir un chat comme un être vivant. Dans un cas comme dans l'autre, cela s'impose à moi. Il y a quelque chose dans l'image et dans l'objet qui me force (car je peux difficilement m'y opposer) à lui appliquer une ou plusieurs interprétations particulières.

5.3. Propriétés charmantes et suspectes

Nous avons vu que de nombreux chercheurs en biologie synthétique sont convaincus que la vie est une propriété émergente de la matière et que la reproduction en laboratoire des mécanismes qui lui ont donné naissance il y a près de quatre milliards d'années nous permettra enfin de la comprendre et de la recréer. Un certain nombre d'entre eux espèrent sans doute observer un jour au fond de leurs éprouvettes la transition de l'inerte au vivant. Fût-elle diffuse ou graduelle, la distinction entre l'un et l'autre existe bel et bien dans la nature, pensent-ils. Notre compréhension de ce qu'est la vie, dit par exemple Bedau, s'incrémentera à mesure que nous en apprendrons davantage et que nous arriverons à créer des choses de plus en plus vivantes (voir B. Holmes, 2005). La vie et la non-vie,

ajoute-t-il, ne sont pas séparées par une ligne claire et distincte mais par une zone grise, et il imagine que l'on pourra mesurer le degré de luminosité de ce gris.

Qu'arriverait-il si la distinction n'existait pas dans la nature mais plutôt dans l'œil de l'observateur, et si la transition n'avait pas lieu entre deux états d'organisation de la matière mais entre deux formes de description de celle-ci ? Peut-être cela ne ferait-il aucune différence, si ce n'est que certains continueraient en vain à chercher dans l'objet lui-même la nature de la vie et le mécanisme de transition. En d'autres termes : le « mystère » de la vie, pour certains, resterait entier.

Cela fait penser au problème de la mesure en mécanique quantique : avant d'être observées, les variables d'un objet quantique sont décrites par un formalisme probabiliste ; après leur observation, les mêmes variables ont une valeur définie ; la difficulté consiste à comprendre comment se fait le passage de l'un à l'autre au moment de la mesure. Un certain nombre de théories ont cherché à expliquer le mécanisme physique qui en est responsable, sans succès (Espagnat, 1994). D'autres, dont le prix Nobel de physique Eugene Wigner, ont même invoqué l'influence mystérieuse de la conscience de l'observateur sur l'objet observé. Des expériences très sérieuses, dites à choix retardé, ont même été mises en place pour « tester » — en partie — cette hypothèse. À l'heure actuelle le problème de la mesure reste une énigme. À moins de considérer que ce qui change au moment de la mesure n'est pas l'état physique de l'objet mais l'état de notre connaissance et le langage employé pour la décrire (Bitbol, 1996, 2000). Selon cette interprétation, la mécanique quantique ne porterait pas directement sur la nature mais sur la connaissance que nous en avons.

Une interprétation analogue dans le domaine de la biologie reviendrait à dire en quelque sorte que l'attribution de vie à un objet ne fait pas référence à un état propre de cet objet mais à la relation (de connaissance ou autre) que nous entretenons avec lui. La vie, alors, ne serait pas une propriété émergente de l'organisation de la matière — ou du moins pas une propriété indépendante de notre propre existence en tant qu'observateurs (?) de celle-ci. De quel type de propriété pourrait-il donc s'agir ? Les observables de la mécanique quantique (ce que l'on appelle classiquement des propriétés ou des variables : la position, la quantité de mouvement, etc.) ne sont définies que relativement à un observateur et à un montage expérimental. Ce sont, suivant l'interprétation que nous venons d'évoquer, des qualités secondes. La vie, suivant l'argument que nous sommes en train de développer, pourrait aussi être une qualité seconde, mais pas forcément du même type.

Daniel Dennett (1993, p. 470) établit une distinction intéressante parmi les qualités secondes entre ce qu'il appelle les propriétés charmantes et les propriétés suspectes. Une personne n'ayant jamais été vue (parce qu'elle a grandi sur une île déserte) pourrait parfaitement être

charmante sans que personne ne l'ai jamais considérée comme telle. En revanche, elle ne pourrait pas être suspecte sans avoir été suspectée (de quelque chose) par quelqu'un. Si elle est charmante, c'est parce qu'elle a le pouvoir dispositionnel d'affecter de façon déterminée les observateurs normaux d'une certaine classe — bien qu'elle n'ait jamais eu l'occasion de le faire. En effet, on ne peut pas définir les qualités charmantes indépendamment des inclinations, susceptibilités ou dispositions d'une classe d'observateurs, et cela n'aurait aucun sens de parler de l'existence de ces propriétés sans évoquer celle des classes d'observateurs pertinentes. Cela étant, une fois définies, leur détermination ne dépend pas d'une observation effective. Les observables de la mécanique quantique ne peuvent donc pas être charmantes ; elles sont plutôt suspectes.

La vie est-elle charmante ou suspecte ? La réponse à cette question n'est pas évidente. D'un côté, il semble que la distinction vivant–inerte ne soit pas définie indépendamment de nous et qu'il n'y ait pas de réponse préétablie à la question de savoir si une entité est vivante ou pas. Qui plus est, nous avons dit que la définition de la vie et l'extension de ce concept sont sujettes à négociation et à révision et qu'elles varient selon la culture. D'un autre côté, il semble que le statut des objets qui nous sont familiers soit intuitivement évidente et que, d'une façon ou d'une autre, elle s'impose à nous. Nous ne choisissons pas de considérer certains objets (les chats, les chiens, les personnes) comme des êtres vivants et les variations interculturelles à leur propos sont sans doute assez faibles. De plus, il est probable que cette capacité intuitive soit liée à la nécessité biologique d'identifier rapidement les proies et les prédateurs. On pourrait donc penser que les objets familiers sont charmants (leur statut vital s'impose de façon évidente) tandis que les entités nouvelles sont suspectes. Cette conclusion est confirmée par des études de psychologie cognitive qui montrent que nous avons effectivement une tendance innée à catégoriser certains objets comme vivants ou, plus généralement, comme des agents. Nous y reviendrons plus longuement au CHAP. VIII.

6. Catégorisation et connaissance

La couleur est une propriété charmante dont les catégories se comportent de façon similaire à la catégorie des êtres vivants. Les couleurs ne sont pas définies indépendamment de notre capacité à les percevoir et il n'y a pas de frontière tranchée entre les unes et les autres, de sorte qu'il n'y a pas non plus de réponse préétablie à la question de savoir si une nuance chromatique donnée appartient à une catégorie ou à une autre. De plus, le nombre et l'extension des catégories de couleur varie en fonction de la culture, bien que certaines d'entre elles (blanc, noir, rouge, vert, bleu) soient communes à presque toutes les cultures et que certains tons apparaissent de façon évidente comme appartenant ou n'appartenant pas à l'une de ces catégories.

6.1. Théorie des prototypes

Dans les années 1970, des recherches de psychologie portant sur la catégorisation ont montré que nous regroupons souvent les objets et les propriétés sur la base de jugements de similarité plutôt que sur la base de définitions et de critères d'appartenance. Ces études, dont les premières ont été menées par Eleanor Rosch (1973, 1975) sur les couleurs, ont conduit au développement de la théorie des prototypes selon laquelle certains exemplaires d'une catégorie donnée sont meilleurs que d'autres. Tandis que le découpage du spectre en portions correspondant aux noms de couleurs élémentaires est largement variable selon les cultures et parmi les membres d'une même culture, il y a au contraire une concordance générale lorsqu'il s'agit de désigner le meilleur exemple d'un nom de couleur élémentaire. Les catégories de couleur ne sont donc pas uniformes : elles ont un membre central prototypique et elles s'étendent ensuite par rayonnement sans limite préétablie. De la même façon, dans le domaine du vivant, les sujets considèrent que les chats sont un exemplaire plus typique de la catégorie « animal » que les tortues et les amibes. Ainsi, les concepts n'ont pas une structure définitionnelle mais probabiliste au sens où quelque chose tombe sous un concept lorsqu'il possède un nombre suffisant d'attributs associés à ce concept (Margolis & Laurence, 2014).

Selon Rosch (1978), la catégorisation repose sur deux principes. Le *principe d'économie cognitive* affirme que la tâche des systèmes de catégories est d'apporter un maximum d'informations sur l'environnement avec un minimum d'effort cognitif. Il s'agit de trouver un équilibre entre la tendance à former le plus grand nombre de catégories possible, car une catégorie plus spécifique contient plus d'information sur ses membres, et la tendance à limiter la quantité d'information à traiter cognitivement, en ignorant les informations non pertinentes, c'est-à-dire en réduisant le nombre de catégories. Le *principe d'exploitation de la structure du monde*

perçu, quant à lui, affirme que le monde perçu se présente de manière structurée et pas sous forme d'attributs arbitraires ou imprédictibles. En effet, on rencontre certaines combinaisons d'attributs plus souvent que d'autres (par exemple, les créatures ailées ont aussi tendance à être couvertes de plumes plutôt que de fourrure) et en formant nos catégories nous exploitons cette structure corrélationnelle (Pacherie, s. d.). Par conséquent, on peut obtenir le maximum d'informations avec le minimum d'effort cognitif quand les catégories s'ajustent à la structure du monde perçu aussi étroitement que possible.

Cela signifie que les catégories, y compris lorsqu'elles portent sur des qualités secondes comme les couleurs, et malgré le fait qu'elles varient d'une culture à une autre, et même parfois d'un individu à un autre, n'en sont pas moins des outils de connaissance du monde extérieur. Les catégories sont l'une des manières que *nous* avons de donner un sens aux choses qui nous entourent, que ce soit pour prédire ou anticiper leur comportement ou pour communiquer avec nos congénères. On cite souvent l'exemple des nombreux mots inuits pour désigner la neige et la glace (Dorais, 2011). Ces mots donnent aux Inuits la capacité d'établir des distinctions très subtiles qui reflètent leur connaissance profonde du milieu dans lequel ils vivent, une connaissance qui n'est pas seulement théorique mais pratique, car liée à l'action : se déplacer, faire un iglou, fondre de l'eau, etc.

La théorie des prototypes de Rosch a ses limitations, mais elles sont instructives. Par exemple, elle fait de l'appartenance catégorielle une questions de degré. Or, certains concepts sont binaires ou dichotomiques, comme le vivant et l'inerte ou la vie et la mort. On pourrait donc considérer que cette théorie ne leur est pas applicable. Cependant, de nombreux chercheurs en biologie synthétique s'expriment comme si ces concepts étaient susceptibles de degrés, car ils se trouvent face à des objets qui d'un côté ressemblent beaucoup aux êtres vivants typiques dans la mesure où ils partagent avec eux un certain nombre de caractéristiques centrales (métabolisme, croissance, reproduction, etc.) mais qui d'un autre côté sont suffisamment différents pour qu'on ne les reconnaisse pas immédiatement comme tels, comme ces particules colloïdales en suspension qui s'auto-assemblent en structures cristallines sous l'effet de la lumière (Palacci et al., 2013). S'il est vrai que la vie et le vivant ont été jusqu'à présent des catégories binaires, c'est peut-être parce que l'on ne connaissait pas encore beaucoup d'objets situés dans cet entre-deux, dans cette zone grise qui sépare le vivant de l'inerte.

Une autre limitation est le défaut de contraintes. C'est-à-dire que la théorie des prototypes ne donne pas de définition ni de pondération des traits jugés pertinents pour établir des relations de similarité. Or, selon les propriétés que l'on considère et l'importance qu'on leur accorde, la similarité entre deux objets sera plus ou moins grande. Mais c'est précisément l'un des problèmes que l'on observe avec les tentatives de définition de la

vie : chacune accorde une importance différente aux attributs et aux propriétés des êtres vivants terrestres et nous ne savons pas comment évaluer leur pertinence indépendamment des définitions elles-mêmes ; c'est-à-dire que chaque définition constitue une catégorie différente à partir d'une ou de plusieurs propriétés centrales et nous ne savons pas sur quelle base choisir l'une des catégories ainsi constituées au détriment des autres.

Une troisième limitation est le défaut de structure. La théorie des prototypes fournit des listes d'attributs, mais elle ne détaille pas la structure des concepts. Or, une représentation des concepts sous forme d'une liste de traits ne permet pas de rendre compte pleinement des relations intra-conceptuelles et inter-conceptuelles. En effet, dit Elisabeth Pacherie :

> [L]es concepts ne sont pas simplement des sommes de propriétés. Un oiseau n'est pas un assemblage quelconque de plumes, d'ailes et de bec. Pour qu'une entité soit véritablement un oiseau, il faut que ces propriétés s'ordonnent en une « structure d'oiseau ». Pour définir ce qu'est une telle structure, il faut faire intervenir des propriétés relationnelles et ne pas se cantonner à des listes d'attributs. (Pacherie, s. d.)

De même, un être vivant n'est pas simplement un ensemble de propriétés ou une liste d'attributs ; il a une structure très particulière dont il faut rendre compte. Comme dit Michel Morange (2007, p. 69): « *La définition de la vie n'est pas à chercher dans une ou quelques caractéristiques qui lui seraient propres, mais dans la réunion et le couplage de ces caractéristiques.* » C'est ce que fait Bedau (2007) en invoquant un principe supérieur, l'évolution indéfinie, pour rendre compte des trois propriétés fonctionnelles qui d'après lui caractérisent le vivant, à savoir une identité (*un conteneur*), l'utilisation d'énergie libre (*un métabolisme*) et un support informationnel (*des gènes*). Sauf que la théorie autopoïétique de Varela (1974) et l'approche physicienne de Morowitz & Smith (2007) proposent également une structure particulière du vivant et un principe général rendant compte de cette structure. Sans compter une approche historique comme celle de Millikan[122].

122 Les trois propriétés de Bedau sont des conditions de possibilité de l'évolution, tandis que dans une approche comme celle de Millikan l'unité n'est pas structurelle ni fonctionnelle mais historique : peu importent les propriétés du vivant et peu importe que celles-ci varient avec le temps, ce qui compte c'est la continuité historique.

6.2. Théorie de la théorie

Étant donné les limitations de la théorie des prototypes, d'autres approches ont été développées plus récemment (voir Margolis & Laurence, 2014). L'une d'elles est la « théorie de la théorie » qui conçoit le processus de catégorisation à la manière du raisonnement scientifique et les concepts comme des théories de la catégorie. Selon cette approche, les concepts encodent de l'information sur les relations structurelles, causales ou fonctionnelles entre les traits ou propriétés qui leur sont associées. Nos croyances sur ces relations explicatives influent sur la catégorisation et elles expliquent le choix des traits représentés dans les concepts : parmi toutes les propriétés d'un objet, nous choisissons celles qui sont rendues saillantes par nos intérêts et connaissances et qui forment un tout cohérent interdépendant (Pacherie, s. d.). Cela implique que nos concepts évoluent à mesure que nos croyances changent, que ce soit avec l'âge (maturité intellectuelle) ou avec la culture (développement scientifique). C'est ainsi que la vie peut d'abord être associée avec le mouvement, puis avec le mouvement autonome ou la respiration, ou encore avec la croissance et la reproduction, et gagner ensuite en complexité avec des propriétés telles que le métabolisme et l'évolution darwinienne. Cela implique aussi que les concepts sont propres à chacun et qu'il est improbable que deux personnes partagent exactement les mêmes concepts. C'est l'une des limitations de cette théorie, mais c'est aussi peut-être, si elle est vraie, l'une des raisons pour lesquelles les scientifiques ne se mettent pas d'accord sur la définition de la vie.

Un autre aspect de la théorie est l'essentialisme : nous attribuons des propriétés cachés aux membres de certaines catégories ontologiques (le vivant) comme s'ils avaient une essence, et les concepts relatifs à ces domaines rendent compte de cette essence. C'est pourquoi nous refusons certains changements de catégorie : un robot ne peut pas devenir un chien, ni devenir intelligent ou vivant, parce que ces concepts impliquent certaines propriétés cachées que les membres des catégories correspondantes ont en commun et que les artefacts n'ont pas (Inagaki & Hatano, 2006; Keil, 2013; Kelemen & Carey, 2007). C'est peut-être ce qui se trouve derrière le vitalisme et l'animisme : les êtres vivants ont une âme, une force vitale, un élan ou quelque chose d'autre qui les distingue essentiellement de la matière inerte. Mais cela pourrait également être à l'œuvre dans les débats actuels sur la définition de la vie. En effet, l'absence de consensus pourrait s'expliquer par le fait que les définitions explicites font référence à des propriétés observables que l'on rencontre aussi en dehors du domaine de la vie, ce qui veut dire qu'elles ne saisissent pas l'essence cachée que nous attribuons intuitivement aux êtres vivants. Et si cette essence n'est pas dans les objets mais seulement dans nos concepts, alors, comme disait Haldane, nous ne pourrons jamais donner de réponse complète à la question : « qu'est-ce que la vie ? »

6.3. Propriétés charmantes et connaissance du monde perçu

Pour en revenir aux propriétés charmantes de Dennett et aux deux principes de Rosch, il faut chercher la structure des êtres vivants non pas seulement dans les objets eux-mêmes mais aussi dans nos structures cognitives. Les êtres vivants sont ce que nous considérons comme tel (l'unité est dans le regard, pas dans les choses), mais la catégorie n'est pas arbitraire (car autrement ce ne serait pas une catégorie scientifique) ni subjective (puisqu'elle ne reflète pas des préférences personnelles). D'un côté, notre prédisposition à voir certains objets comme vivants est peut-être le résultat de pressions sélectives auxquelles ont été soumis nos ancêtres. D'un autre côté, si cette catégorie est le résultat d'une pression sélective, c'est vraisemblablement parce qu'elle nous apporte une information pertinente sur le monde. Il faudrait donc savoir ce qui, dans le monde, correspond à nos structures cognitives. En d'autres termes, si la vie est une propriété charmante, il faudrait savoir ce qui, dans les objets, les dispose à être perçus comme des êtres vivants. Nous y reviendrons au CHAP. VIII et dans la cinquième partie.

Prenons l'exemple des mauvaises odeurs. La distinction entre bonnes et mauvaises odeurs n'a aucune signification physique : il n'y a rien qui fasse qu'une odeur soit intrinsèquement agréable ou désagréable, car c'est seulement *pour nous* qu'elles le sont. Pourtant, il y a bien quelque chose qui dispose certaines substances à sentir mauvais : c'est leur composition chimique. Les composés à base de souffre, par exemple, sont souvent toxiques pour nous, et l'évolution biologique nous a prédisposés à ne pas aimer les odeurs de produits toxiques, d'où l'expérience subjective désagréable. Autrement dit, la valeur hédonique d'une odeur est en partie prédéterminée par la structure de la molécule odorante avant d'être modulée par l'expérience et la culture (Kermen et al., 2016).

En ce qui concerne les êtres vivants, ce qui les dispose à être perçus comme tels est vraisemblablement, d'une part, la complexité de leur organisation (d'où l'analogie de la montre de William Paley) et, d'autre part, leur autonomie (mouvement autonome, auto-production, auto-réplication, homéostasie, réponse à des stimuli, etc.), les deux choses étant liées. Or, ces deux aspects sont aussi sans doute ce qui nous prédispose à leur attribuer des fins. Il y a donc probablement un lien entre la disposition à classer un objet dans la catégorie des êtres vivants et la disposition à lui attribuer des fins.

Si la vie est une propriété charmante, si elle ne constitue pas un genre naturel mais un genre nominal, si la frontière entre le vivant et l'inerte n'est pas dans la nature mais dans le regard, et s'il n'y a pas une essence ni une structure interne ou des propriétés communes qui puissent en rendre compte, cela ne veut pas dire que le concept, la catégorie, la distinction et la définition soient pour autant arbitraires, car s'il est vrai

que la catégorisation est une source de connaissances, et si cette connaissance repose sur l'exploitation de la structure du monde perçu, alors la catégorisation du vivant, aussi subjective qu'elle puisse paraître, n'en constitue pas moins une connaissance à part entière du monde vivant, une connaissance qui se traduit par exemple en termes de succès prédictifs et de généralisations inductives[123].

De la même manière, on peut penser que l'attribution de fins à certains objets et pas à d'autres est l'expression d'une connaissance authentique du monde extérieur et pas seulement le vestige de modes d'explication pré-scientifiques. Reste à savoir ce qui justifie l'attribution de fins à ces objets, au-delà des stratégies naturalistes que nous avons déjà précédemment, et comment l'attribution de fins est liée à l'attribution de fonctions.

123 On pourrait évidemment discuter du type de connaissance que cela suppose, mais nous n'allons entrer ici dans ce débat.

Conclusions de la deuxième partie

Les fonctions et le vivant semblent être intimement liés, puisque tous les porteurs de fonctions (éléments organiques, comportements, artefacts) appartiennent ou sont le produit, ou sont associés (virus) à des êtres vivants. On pouvait donc supposer que la connaissance scientifique du vivant nous aiderait à mieux comprendre les fonctions. Toutefois, on constate qu'il n'y a pas davantage de consensus à propos de l'un qu'il n'y en a à propos de l'autre. Quand on se penche sur les recherches en biologie synthétique et sur les origines de la vie, on remarque que les doutes et les difficultés sont similaires à celles rencontrées à propos des fonctions et que, dans un cas comme dans l'autre, les spécialistes sont partagés entre les mêmes types d'approches : systémiques, darwiniennes et mixtes.

Comme pour les fonctions, ces différentes approches ne situent pas la frontière entre le vivant et le non-vivant au même endroit, et elles ont chacune de leur côté du mal à établir une distinction non-ambigüe entre les deux domaines. On remarque que les critères employés dans ce débat sont également les mêmes que ceux du débat sur les fonctions, à savoir la complexité et la sélection naturelle, lesquels soulèvent pratiquement les mêmes objections et les mêmes problèmes dans les deux cas.

Ces difficultés nous amènent à nous demander si la vie et le vivant existent vraiment, s'ils correspondent à un genre naturel ou à une grappe naturelle, ou à un ensemble d'objets naturellement distincts dont la biologie aurait pour mission de découvrir la nature, les limites, les propriétés et les mécanismes. Notre réponse est négative : la vie est une catégorie humaine — ou plutôt une classe humaine (human kind) — façonnée à la fois par l'intuition, la culture et la science. Certes, toutes les catégories sont humaines dans la mesure où elles sont le produit de nos pratiques classificatoires, car pour rendre le monde intelligible à notre entendement, nous le découpons en différentes parties, mais on conçoit habituellement que le découpage peut respecter les articulations naturelles des parties ou au contraire les mutiler, comme le ferait un mauvais boucher. Ce que nous disons ici, c'est qu'il n'y a pas d'articulation naturelle entre la vie et la non-vie, pas plus qu'il n'y a une essence ni une structure interne commune aux êtres vivants. Nous pouvons donc choisir où trancher, mais tous les choix ne se valent pas. En fonction de nos inté-

rêts et de nos pratiques épistémiques, certaines découpes s'avèrent être meilleures que d'autres.

Il en est de même des *continents* en géologie et des *planètes* en astronomie. Ces dernières ne sont pas un genre ou une classe naturelle, ni un ensemble de corps naturellement distincts des non-planètes, mais une classe humaine. La définition formulée en 2006 par l'Union Astronomique Internationale est stipulative, conventionnelle et pragmatique ; elle ne prétend pas dire ce que *sont* les planètes, car celles-ci sont avant tout un concept culturel, mais se contente de restreindre l'usage du terme « planète » dans le Système Solaire à certains types de corps, afin que tous les astronomes l'emploient de la même façon. Les planètes extrasolaires, quant à elles, ne sont pas définies. Cela ne veut pas dire que la distinction soit arbitraire ; elle ne l'est pas et s'appuie sur des critères scientifiquement pertinents et à la pointe des connaissances actuelles.

Si les êtres vivants n'étaient pas un genre ni une grappe naturelle, s'ils n'étaient pas naturellement distincts des objets inertes, ou bien si l'appartenance d'un objet au domaine du vivant n'était pas déterminée indépendamment de nous, alors la scientificité de la biologie ne serait pas remise en question, pas plus que ne l'est celle de l'astronomie.

Les difficultés rencontrées pour identifier les frontières du vivant sont dues au fait que nous cherchons une transition entre deux états de la matière là où il y a en réalité une transition du regard. Autrement dit, ce ne sont pas les objets eux-mêmes qui changent, mais notre façon de les décrire et de les étudier. Depuis cette perspective, on pourrait concevoir le passage du regard physique au regard biologique à la manière de celui qui se produit face à une image ambiguë comme la figure de Jastrow. Cela impliquerait que la vie est une qualité seconde, parmi lesquelles Daniel Dennett distingue les propriétés « charmantes » et « suspectes ».

La catégorisation des propriétés charmantes implique une forme de connaissance du monde extérieur. En effet, même si la vie et le vivant ne sont pas une catégorie naturelle, le fait de classer un objet parmi les êtres vivants n'est pas d'arbitraire. Il est l'expression d'une relation entre nos structures cognitives et les structures du monde perçu, lesquelles sont souvent la manifestation de structures plus profondes. C'est cette relation qui rend possibles les généralisations inductives, les prédictions réussies et les explications scientifiques. Il y a donc là une connaissance objective qui ne requiert pas un engagement ontologique vis-à-vis de l'objet connu.

Si les fonctions et le vivant sont intimement liés, alors il n'y a pas non plus de distinction naturelle, indépendante de nous, entre ce qui est fonctionnel et ce qui ne l'est pas, ni une nature commune ou un ensemble de propriétés partagées par tous les porteurs de fonctions. Malgré tout, les attributions fonctionnelles ne sont pas arbitraires, puisqu'elles n'ont pas toutes la même valeur explicative. Certaines nous aident effectivement à comprendre les phénomènes biologiques, tandis que d'autres ne le font pas ou dans une moindre mesure. Or, ce n'est pas

nous qui choisissons que certaines attributions soient meilleures que d'autres du point de vue scientifique. Il y a donc manifestement une part d'objectivité dans cette connaissance. Mais cela ne veut pas dire non plus que la fonction des traits biologiques soit écrite quelque part dans le Livre de la Nature, indépendamment de la connaissance que nous en avons.

Il n'y a peut-être pas de réponse naturelle, prédéterminée, attendant d'être découverte, à des questions comme « Un cœur malformé a-t-il la fonction de pomper le sang alors qu'il en est incapable ? », « Les organes des doubles accidentels sont-ils fonctionnels ? » ou « Les manchots ont-ils des ailes ou des nageoires ? ». Comme pour le statut planétaire de Pluton, la question n'est peut-être pas de savoir si un trait possède ou pas une fonction, mais de savoir dans quelle mesure la lui attribuer est à la fois pertinent et utile d'un point de vue scientifique, c'est-à-dire dans quelle mesure cette attribution est un acte et une source de connaissance.

Troisième partie :

CARACTÉRISATION GÉNÉRALE DES FONCTIONS

I have stressed the importance of the use of such concepts as biological means and ends because I want it clearly understood that I think that such a conceptual framework is the essence of the science of biology.

George C. Williams, *Adaptation and natural selection (1966, p. 11)*

Functions are intimately related to ends of certain kinds, and the items to wich functions are attributed are means to those ends.

Peter Achinstein, « Function statements » (1977, p. 359)

[T]he essential feature of a functional relationship is that of a *means-end* relation. For a structure x to have a function y is, essentially, for x to do y in a system S and for y to lead to the fact that the system is able to output a value O. The output value O will either be a goal-state of S or causally contribute to S's attaining a goal-state.

Frederick Adams,
« A goal-state theory of function attributions » (1979, p. 494)

Introduction de la troisième partie

L'objectif de cette troisième partie est de formuler une explication philosophique du concept de fonction. Nous entendons par là un type d'analyse qui consiste à montrer qu'un concept vague ou problématique peut être avantageusement remplacé par un autre plus précis et qui joue le même rôle, mais qui ne pose pas les mêmes problèmes. Contrairement à l'analyse conceptuelle, il ne s'agit pas de capturer la signification d'un concept, ni de décrire ce que les gens ont à l'esprit quand ils l'emploient, ni de rendre manifestes ses conditions d'application. La méthode que nous allons suivre est celle que propose Schwartz (2004) à la suite de Carnap (1950) et de Quine (1960). Ce dernier parle de substituer un concept confus par un autre meilleur :

> « We do not claim synonymy. We do not claim to make clear and explicit what the users of the unclear expression had unconsciously in mind all along. We do not expose hidden meanings, as the words "analysis" and "explication" would suggest; we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions. Beyond those conditions of partial agreement, dictated by our interests and purposes, any traits of the explicans come under the head of "don't-cares". » (Quine, 1960, p. 258-9)

Puisque les fonctions biologiques posent principalement un problème de légitimité et d'acceptabilité scientifique, nous allons essayer de montrer qu'il n'est pas nécessaire d'analyser les fonctions en termes de causes pour les rendre scientifiquement acceptables. Pour y arriver, nous allons proposer tout d'abord une interprétation et une caractérisation du concept de fonction en termes téléologiques. Ensuite, dans la quatrième partie, nous tâcherons de justifier ce recours à la téléologie dans un cadre scientifique. Finalement, dans la cinquième partie, nous développerons davantage notre explication philosophique du concept de fonction en relation avec d'autres approches non causales.

Par ailleurs, nous verrons que notre explication des fonctions évite les défauts des conceptions étiologiques et systémiques, et qu'elle les subsume, c'est-à-dire que la définition que nous proposons est suffisamment générale pour être compatible avec la plupart des définitions précédentes entendues comme des cas particuliers, et assez abstraite pour dépasser la distinction entre les fonctions biologiques et techniques.

Étant donné que les conceptions étiologiques et systémiques ne tombent pas toujours d'accord sur le fait qu'un trait donné possède ou pas une fonction, il peut s'avérer nécessaire de faire un choix en faveur de l'une ou de l'autre. Par exemple : un cœur possède-t-il la fonction de pomper le sang s'il en est incapable ? Nous pensons que oui, contrairement à Cummins et à Davies. Cela ne veut pas dire que l'une des réponses soit vraie et l'autre fausse. Notre explication du concept de fonction ne prétend pas être vraie ni correcte, mais seulement satisfaisante :

> « if a solution for a problem of explication is proposed, we cannot decide in an exact way whether the solution is right or wrong. Strictly speaking, the question whether the solution is right or wrong makes no good sense because there is no clear-cut answer. The question should rather be wether the proposed solution is satisfactory, whether it is more satisfactory than another one, and the like. » (Carnap, 1950, p. 4)

Nous partageons aussi avec Millikan (2002, p. 122) l'idée qu'il n'existe pas une unique manière correcte d'employer le concept de fonction et qu'il ne faut pas nécessairement chercher à lui donner une définition très précise, non seulement parce que les distinctions naturelles sont parfois vagues et indéterminées, mais aussi parce que, comme nous l'avons suggéré précédemment, la distinction n'est peut-être pas naturelle mais humaine. Il pourrait donc y avoir autant d'acceptions et de définitions correctes de ce concept qu'il y a d'usages sanctionnés par la communauté scientifique ou philosophique. Dès lors, nous ne prétendons pas découvrir un ensemble magique de conditions nécessaires et suffisantes qui permettrait de couvrir tous les cas possibles, d'éviter tous les contre-exemples et de répondre à toutes les questions.

La recherche d'une définition générale n'exclut pas le pluralisme. Plusieurs définitions incompatibles sont en effet envisageables, et c'est alors aux biologistes de choisir laquelle leur convient le mieux pour rendre compte du sens littéral des attributions fonctionnelles :

> « In this case, biologists would be free to choose either account to explain the literal meaning of their terms. As long as we are looking for a way to replace traditional, problematic concepts with new, useful ones, *two* options are as good (if not better) than one. » (Schwartz, 2004, p. 145)

De plus, l'explication philosophique n'exige pas des biologistes qu'ils adoptent et qu'ils suivent au pied de la lettre la nouvelle définition. Ils pourraient, dit Schwartz, continuer à employer le mot « fonction » de manière ambigüe et même contradictoire — ce qui présente peut-être des avantages d'un point de vue heuristique ou métaphorique. Cependant, pour qu'une attribution fonctionnelle puisse être prise littéralement, ils devraient être disposés à adopter une définition précise et montrer que l'attribution en question en satisfait les conditions.

L'explication philosophique ne se prononce donc pas quant au caractère téléologique du concept de fonction tel qu'il est employé par les biologistes. Elle se contente de dire que l'on peut attribuer des fonctions sans avoir recours à des notions problématiques, car il est possible de formuler une définition qui les évite. Or, ce qui est problématique avec les fonctions et la téléologie en général, ce sont les causes finales, la causalité rétrograde, la prédétermination de l'avenir, etc. La plupart des auteurs étudiés jusqu'à maintenant proposent donc des définitions du concept de fonction formulées en termes de causalité efficiente. La nôtre, au contraire, est formulée en termes téléologiques, car la téléologie ne présente pas nécessairement les défauts qu'on lui impute, comme nous le verrons par la suite.

CHAPITRE VI

La fonction comme contribution à une fin

En première approximation, nous dirons que les fonctions d'un item sont à la fois relatives à un but ou à une fin et relatives à un système contenant. Plus précisément, nous défendrons l'idée que *la fonction d'un item est sa contribution à la fin du système auquel il appartient*.

Pour expliquer cette idée, nous allons partir de la proposition de Kitcher, examinée dans la première partie, que nous compléterons avec d'autres considérations. Après avoir précisé ce que notre proposition implique concernant la manière de comprendre les explications fonctionnelles, nous aborderons sur trois questions importantes pour la suite, à savoir : le typage fonctionnel, la distinction entre conséquences fonctionnelles et accidentelles, et l'adaptationnisme.

1. Première approximation

On se souvient que chez Kitcher (1993) la fonction d'un item est sa contribution à la réalisation de « ce pour quoi le système a été conçu », contrairement à « ce pour quoi le système a été sélectionné » que l'on trouve chez Neander (1991b). Dans le domaine biologique, ce sont les pressions sélectives de l'environnement auxquelles l'organisme doit répondre qui déterminent ce que l'on peut appeler sa finalité (c'est-à-dire ce pour quoi il a été conçu, intentionnellement ou pas). Autrement dit, un organisme dans un environnement donné doit répondre à des pressions sélectives (sous peine de disparaître) et l'attribution d'une fonction à un trait dépend de son rôle dans la réponse que l'organisme oppose à ces pressions. Plus précisément, après avoir identifié les pressions auxquelles il est soumis, en particulier lors d'un changement dans ses conditions environnementales (comme les lézards sur l'île de Pod Mrčaru), on peut étudier sa réponse à ces pressions (évolution de la

morphologie, apparition de valves cæcales, etc.) et lui appliquer une analyse fonctionnelle à la Cummins pour établir la contribution causale respective des différents traits impliqués dans cette réponse.

Plus généralement, la proposition de Kitcher obéit au principe selon lequel la fonction d'un item est *ce pour quoi* il a été conçu indépendamment de *la manière dont* il a été conçu, donc indépendamment de la nature intentionnelle ou pas de cette conception. Or, la fonction d'un item peut être déterminée par ingénierie inverse du système auquel il appartient. Donc, si l'on connaît ou si l'on peut déduire (de son organisation, de son contexte) la finalité d'un système, alors on peut définir les fonctions de ses parties d'après leur contribution causale — actuelle ou dispositionnelle, réelle ou prétendue — à la réalisation de cette fin, quelles que soient par ailleurs les origines du système (création consciente, sélection naturelle ou autre mécanisme causal) et quel que soit l'état dans lequel il se trouve (vivant ou mort, « en état de marche » ou « cassé »).

L'analyse fonctionnelle telle que l'applique Kitcher, ou du moins telle que nous l'entendons, ne constitue pas une explication causale de la présence d'un item et ne correspond donc pas aux attributions fonctionnelles des approches étiologiques, et cela d'autant moins lorsqu'elles font référence au mécanisme de sélection naturelle. Nonobstant, on peut rapprocher la proposition de Kitcher de la formulation étiologique abstraite de Larry Wright (1973) où la fonction d'un item est entendue comme sa raison d'être plutôt que comme la cause de sa présence : (1) le cœur des vertébrés existe parce qu'il contribue à faire circuler le sang et (2) la circulation du sang chez les vertébrés est une conséquence de la présence du cœur ; mais peu importe que sa présence soit le résultat d'un processus de sélection naturelle (hypothèse scientifique) ou d'une conception intentionnelle (« hypothèse » créationniste), puisque dans les deux cas c'est sa contribution à la circulation du sang qui explique pourquoi les vertébrés en ont un.

Il ne s'agit pas non plus de fonctions « à la Cummins » car l'attribution d'une fonction à un item ne dépend pas de sa disposition effective à contribuer à une capacité présente du système. Il n'est par exemple pas absurde d'appliquer l'analyse fonctionnelle à un organisme fossile pour attribuer à ses parties des fonctions relatives aux capacités qu'il possédait lorsqu'il était vivant bien qu'actuellement ni le système ni ses parties n'aient plus les capacités en question. L'analyse fonctionnelle du système circulatoire chez un dinosaure devrait aboutir aux mêmes conclusions indépendamment de son état vital : le cœur d'un dinosaure mort n'a plus la capacité de faire circuler le sang, mais sa fonction dans l'organisme reste la même. On devrait donc pouvoir conjuguer l'analyse fonctionnelle de Cummins au passé : « Lorsque le dinosaure était vivant, il avait la capacité de faire circuler le sang dans tout l'organisme et la fonction du cœur dans le système circulatoire était de pomper le sang. » On devrait

également pouvoir la conjuguer au conditionnel : « Si le dinosaure était vivant, il aurait la capacité de… et la fonction du cœur serait de… » Or si cette formulation conditionnelle était acceptable, l'analyse fonctionnelle devrait aussi être applicable aux traits qui ne sont pas en mesure de réaliser la fonction qu'on leur prête : « Si le cœur de cet animal n'avait pas été mal formé, il aurait eu la capacité de pomper le sang et ne serait pas mort-né. » Le même raisonnement est valable pour les artefacts, comme nous l'avons déjà indiqué à propos du mécanisme d'Anticythère (voir page 106) : ce n'est pas parce qu'un appareil est cassé qu'on ne peut pas attribuer de fonctions à ses parties.

La proposition de Kitcher s'inscrit dans le cadre d'une intuition plus générale formulée trente ans auparavant par Canfield (1964), Ayala (1970) et d'une certaine manière aussi par Sorabji (1964), à savoir que la fonction d'un item est ce en quoi il est utile au système auquel il appartient. Lorsque le système est un artefact, son utilité est relative à ses créateurs ou utilisateurs[124]. Lorsque le système est un être vivant, l'utilité est relative au système lui-même dans une perspective darwinienne ; c'est-

124 Canfield ne s'occupe dans son article que des fonctions biologiques, ce qui lui vaut une critique de Larry Wright (1973, p. 145) disant que sa formulation est inapplicable aux fonctions conscientes des artefacts. La formulation en question (simplifiée) est la suivante : « A function of *I* (in *S*) is to do *C* means *I* does *C* and that *C* is done is useful to *S*. » (Canfield, 1964, p. 290). Selon Wright, quand on dit que la trotteuse d'une montre est utile pour faciliter la lecture des secondes, soit on considère que le système *S* est la montre, auquel cas il est absurde de dire que la trotteuse est utile pour la montre, soit on considère que le système pour lequel la trotteuse est utile est l'utilisateur de la montre, auquel cas il est absurde de dire que la trotteuse appartient au système *S*. Wright complète sa critique en y ajoutant deux considérations intéressantes. En premier lieu, si l'utilisateur n'a pas besoin de ce degré de précision dans la lecture de l'heure, alors la trotteuse n'a pour lui aucune utilité bien qu'elle ait toujours la fonction d'indiquer les secondes. En second lieu, il existe des machines qui n'ont aucune utilité mais dont les parties ont néanmoins des fonctions dans la mesure où elles ont été conçues pour (*designed for*) faire quelque chose d'inutile. Ces critiques peuvent être répondues, comme nous allons le voir, en disant qu'un item est utile pour le système auquel il appartient dans la mesure où il contribue à la réalisation de la finalité de ce système. Dans le cas des artefacts, la finalité est extrinsèque, de sorte que le système *S* auquel appartient la trotteuse est bien la montre, mais la finalité de cette dernière est fixée par son créateur ou par son utilisateur et c'est relativement à cette finalité que la trotteuse est utile. Peu importe que la trotteuse soit utile ou pas pour l'utilisateur, car sa fonction n'est pas relative à l'utilisateur mais à la finalité de la montre. Et peu importe que le système lui-même ait une quelconque utilité, car même un système inutile (Wright mentionne une machine du MIT dont la seule fonction était de s'éteindre elle-même) a une finalité et c'est relativement à cette finalité que les parties du système ont une utilité.

à-dire que l'item a une fonction dans la mesure où il contribue au succès reproductif de l'organisme auquel il appartient. Il convient de souligner par ailleurs que chez Canfield (1964, p. 287) l'item et le système ne sont pas des particuliers mais des types ou des classes.

En reprenant la formulation de Canfield on dira ainsi que les battements du cœur chez les vertébrés ont la fonction de faire circuler le sang dans la mesure où le cœur réalise effectivement cette activité (ou en a la capacité) et dans la mesure où elle leur est utile (en contribuant à leur succès reproductif)[125]. L'utilité d'une fonction n'est pas quelque chose d'accidentel, comme dans l'exemple de la boucle de ceinturon qui dévie les balles (L. Wright, 1973, p. 147), mais quelque chose qui permet d'expliquer pourquoi les vertébrés ont un cœur, quel que soit par ailleurs le mécanisme causal invoqué par cette explication[126].

Chez Ayala (1970), l'utilité sert de critère pour distinguer les systèmes véritablement téléologiques de ceux qui ne le sont pas. Il distingue dans ces systèmes deux niveaux de finalité : le premier correspond aux fins des parties d'un système, le second correspond à la fin ou au but du système dans son ensemble. La relation moyens–fin est donc double : chez les êtres vivants, de manière générale, chaque trait a simultanément une fin proximale qui est sa fonction et une fin ultime, à laquelle il contribue, qui est le succès reproductif de l'organisme dans son ensemble[127].

La même idée est attribuable à Hempel et à Nagel pour qui le porteur de fonction (l'item) est un moyen en vue d'une fin, laquelle est à son tour un moyen en vue d'une autre fin :

> *X* a la fonction *Y* ssi *X* est un moyen en vue d'une fin relative *Y* qui à son tour est un moyen en vue d'une fin intrinsèque *G* (et le système *S* doit appartenir à un type de systèmes qui ont ou qui peuvent avoir une fin intrinsèque *G*) (McLaughlin, 2001, p. 76).

---

125 Hugh S. Lehman (1965b) formule un contre-exemple auquel répond Canfield (1965).

126 Pour un scientifique, l'explication de la présence de l'organe passe par la théorie synthétique de l'évolution et donc par le mécanisme causal de la sélection naturelle. Pour un créationniste, l'explication passe par la volonté d'un créateur intelligent. Les boucles de ceinturon auraient également la fonction d'arrêter ou de dévier les balles si l'intention correspondante avait été présente lors de leur création ou de leur utilisation.

127 Mayr (1974) s'oppose à ce que les concepts généraux de survie et de succès reproductif soient désignés comme buts de la sélection naturelle. D'après lui, le but d'un processus téléonomique doit être quelque chose de concret et bien déterminé, de sorte que l'évolution darwinienne ne peut être considérée comme téléonomique.

Nous retiendrons cette idée pour préciser notre affirmation initiale selon laquelle la fonction d'un item est sa contribution à la fin du système auquel il appartient. Tout d'abord, une fonction est une fin qui constitue elle-même le moyen d'une autre fin. Ensuite, l'item auquel on attribue une fonction fait partie d'un tout auquel on attribue une fin, laquelle peut à son tour être une fonction relativement à la fin d'un système plus vaste.

D'un côté, nous avons un emboîtement de structures et de processus matériels. De l'autre, un emboîtement de fins. Or, une même fin peut être atteinte par différents moyens et une même fonction réalisée par différentes structures ou processus. On peut donc retrouver les mêmes relations fonctionnelles sous diverses formes matérielles. C'est la raison pour laquelle, au lieu de considérer un système téléologique donné (artefact ou être vivant) comme étant un ensemble de structures et de processus matériels, nous allons le concevoir à la suite de Wimsatt (2002) comme un ensemble intégré de fonctions implémentées par ces structures et ces processus (voir Fig. 14), c'est-à-dire un ensemble d'opérations[128] que ces structures et processus réalisent ou sont capables de réaliser dans des conditions données.

Jean Gayon exprime cette idée de la manière suivante : « *Ce ne sont pas les structures qui ont une fonction, mais ce qu'elles font, leur effet dans un contexte déterminé. Les fonctions ne sont pas des propriétés monadiques des structures* ». Ainsi, une horloge n'est pas un ensemble de rouages et d'aiguilles mais un ensemble de fonctions dont l'organisation a pour finalité d'indiquer l'heure ; et le fait que ces fonctions soient matérialisées dans

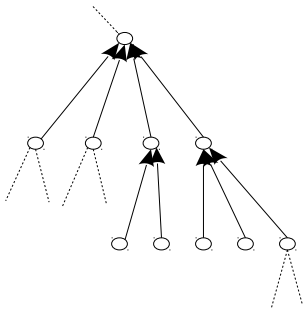

Figure 14: Schéma simplifié d'un système fonctionnel hiérarchique. Les nœuds représentent, non pas des éléments matériels mais des fonctions.
(Source : Wimsatt, 2002)

128 L'idée de la fonction comme étant un ensemble d'opérations est notamment rapportée par Ferrater-Mora (1986, p. 1300) : « *[E]n parte de la filosofía moderna […] interviene cada vez más decisivamente la idea de función como operación o conjunto de operaciones que determinan lo que una realidad es o que permiten comprender esa realidad.* »

une horloge concrète par tels et tels mécanismes matériels est tout à fait contingent car d'autres mécanismes pourraient remplir les mêmes fonctions[129]. L'analyse fonctionnelle d'un système téléologique se situe donc à un niveau d'abstraction plus élevé que l'analyse de son organisation matérielle.

Une autre différence importante entre les approches précédentes et celle que nous proposons ici tient à l'attitude adoptée vis-à-vis du langage téléologique de la biologie. Tandis que la plupart des auteurs cherchent à réduire, traduire ou expliciter les énoncés téléologiques en termes moins compromettants[130], nous affirmons au contraire que les attributions de fins sont tout à fait légitimes, acceptables et parfois même indispensables. Ainsi, contrairement aux approches étiologique et dispositionnelle, nous n'analyserons pas les fonctions en termes d'effets ou de capacités mais en termes de moyens et de fins. Et contrairement à l'approche valorative défendue notamment par Woodfield (1976) et par Bedau (1992b), nous n'analyserons pas les fins seulement en termes de bénéfices, car une fin n'est pas nécessairement bonne ni considérée comme telle.

2. Explications et niveaux d'abstraction

Notre proposition implique une inversion temporelle par rapport aux explications causales. L'existence ou la présence d'un trait biologique s'explique habituellement en faisant référence au travail de la sélection naturelle sur les « ancêtres » de ce trait, c'est-à-dire à une série d'événements situés dans son passé. De manière générale, les définitions étiologiques du concept de fonction expliquent l'existence ou la présence d'un item x en faisant référence à un ou plusieurs items x' antérieurs et à l'histoire causale qui les lie (Fig. 15a). Les définitions dispositionnelles, quant à elles, optent pour une stratégie différente car les fonctions n'y sont pas tournées vers le passé mais vers le présent ou vers le futur : au lieu de porter sur l'item fonctionnel x, leurs explications portent sur les capacités du système S auquel cet item contribue, de telle manière qu'elles aussi soient rétrogrades (Fig. 15b). Dans un cas comme dans l'autre, l'*explanans* est chronologiquement antérieur à l'*explanandum*. Au contraire, dans l'approche que nous proposons, l'explication est tournée vers le futur : elle porte sur l'item fonctionnel x et fait référence à la finalité du système $F_{(S)}$ auquel il appartient (Fig. 15c). Bien entendu, il ne

129 Comme dit Wimsatt (2002, p. 179) : « *No other features of the objets are relevant other than the fact that they do the same things under certain conditions—which is to say that it is their behavior that is important* ».

130 D'après Grene et Depew (2004, p. 316), c'est également ce que font Larry Wright et les partisans d'une conception étiologique basée sur la sélection naturelle.

s'agit pas d'une explication causale — puisque c'est x qui contribue causalement à $F_{(S)}$ et pas l'inverse — mais d'une explication instrumentale. Attribuer une finalité à un système permet en effet de justifier du point de vue de la rationalité instrumentale les moyens mis en œuvre pour sa réalisation. C'est-à-dire que la référence à la finalité d'un système permet de justifier rationnellement l'existence ou la présence des éléments qui le composent.

Réciproquement, l'attribution d'une fonction aux composants d'un système permet d'expliquer ce dernier à partir d'une analyse fonctionnelle à la Cummins où la capacité analysée n'est autre que la finalité de ce système. Or, nous disions plus haut que l'analyse fonctionnelle d'un système téléologique se situe à un niveau d'abstraction plus élevé que l'analyse de son organisation matérielle. Par conséquent, le système dont nous analysons la finalité $F_{(S)}$ n'est pas composé d'éléments matériels x mais d'éléments fonctionnels $f_{(x)}$ (Fig. 15d).

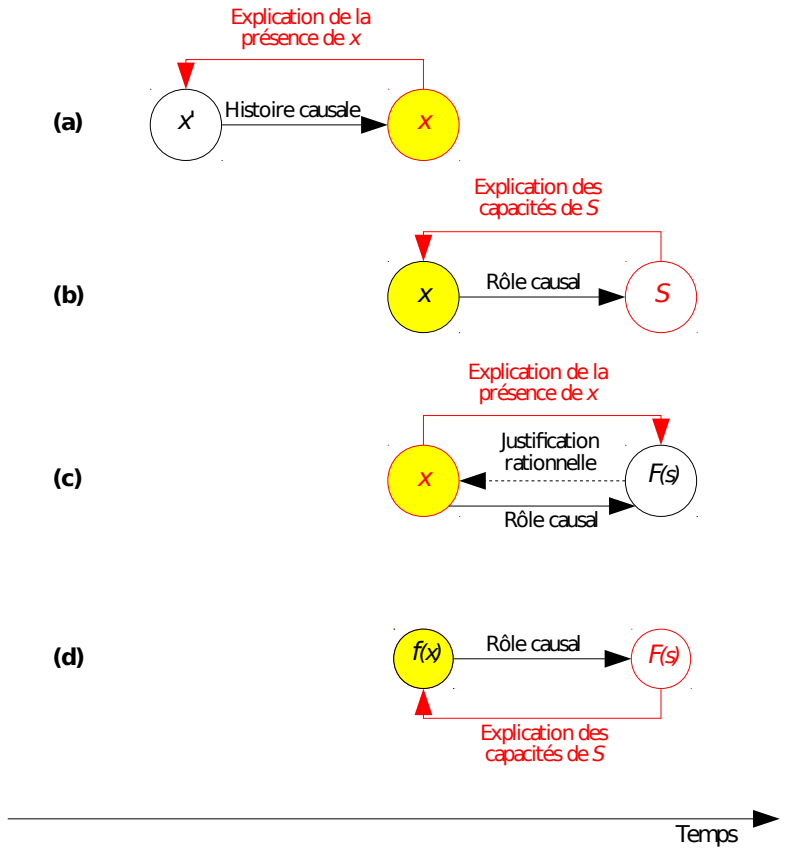

Figure 15: Schémas explicatifs des différentes approches du concept de fonction. Légende : (a) correspond à l'approche étiologique ; (b) correspond à l'approche dispositionnelle ; (c) et (d) correspondent conjointement à l'approche que nous proposons. L'explanandum *est indiqué en rouge, l'*explanans *est indiqué en noir, et le porteur de fonction est indiqué en jaune.*

Par exemple, la finalité du système circulatoire sanguin est d'assurer le transport jusqu'aux cellules de l'oxygène, des nutriments et de certains messagers chimiques, ainsi que la collecte des déchets organiques. Il est composé essentiellement d'une pompe, d'un circuit fermé et de plusieurs échangeurs pour les molécules transportées (Fig. 16). Ces composants désignent ici non pas des éléments matériels mais des activités fonctionnelles (pompage, etc.) pouvant être accomplies par différentes structures et mécanismes causaux, lesquels peuvent à leur tour être analysés fonctionnellement : le cœur est un sous-système du système circulatoire dont la finalité est de pomper le sang et dont les composants sont des activités fonctionnelles pouvant être accomplies par différents mécanismes pouvant à leur tour être analysés fonctionnellement…

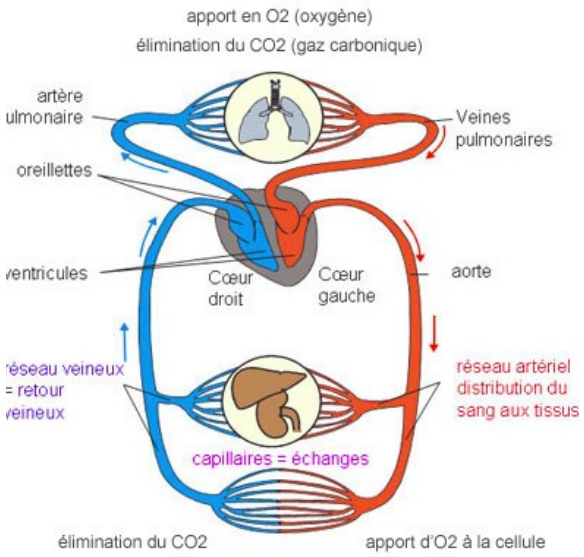

Figure 16: Schéma simplifié du système circulatoire.

Quand, chez un individu concret, on désigne une masse de tissus donnée comme étant un cœur (Fig. 18), on identifie ces tissus à un type d'objets (Fig. 17) définis par leur activité fonctionnelle dans le système circulatoire (Fig. 16). Autrement dit, on établit une correspondance entre un système concret et un système abstrait, et on identifie des tissus réels (ce cœur là) à un élément idéal (la pompe) du schéma fonctionnel correspondant. La même procédure est applicable à n'importe quel artefact technique comme une horloge ou comme le mécanisme d'Anticythère. Dans ce dernier cas, la nature du système n'étant pas connue a priori, les archéologues doivent commencer par formuler des hypothèses quant à sa finalité générale pour ensuite imaginer des schémas fonctionnels s'ajustant aux fragments d'engrenages retrouvés. La fonction de ces fragments ne dépend pas de leur propre capacité à contribuer à la finalité du système, car de toute évidence cette capacité est actuellement nulle, mais des capacités correspondant à leur type. De manière analogue, la fonction de la masse de tissus de la Fig. 18 ne dépend pas de son propre rôle causal dans le système circulatoire de l'organisme concret auquel il appartient, lequel est probablement déjà mort, mais du rôle causal correspondant à son type[131] (Fig. 16 et 17).

131 Voir note 133, p. 237.

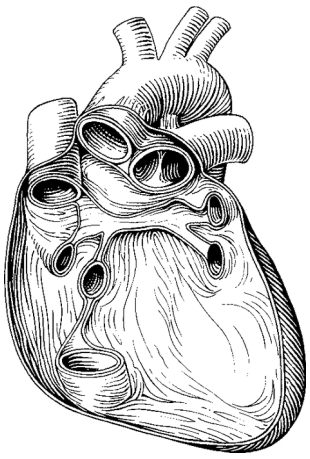

Figure 17: Schéma en coupe du cœur humain. (Source : Feneis & Dauber, 2000)

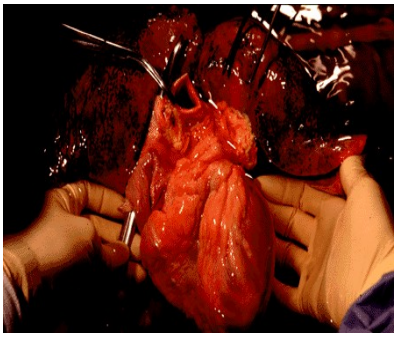

Figure 18: Cœur humain à vif. (Source : Cabrol, Vialle & Guérin-Surville, 2002)

En attribuant une fonction à un item *x*, on donne donc à la fois une explication des capacités du système *S* et une explication de sa présence dans ledit système[132] (Fig. 19e), cette dernière étant parfaitement compatible avec une explication causale (Fig.19f). Mais nous insistons sur le fait que l'explication se déroule à la fois sur deux niveaux : concret et abstrait. Par exemple, pour expliquer l'existence de la masse de tissus de la Fig. 18, il faudrait faire référence à l'histoire causale concrète de l'individu en question. Or cette histoire nous est inconnue. Cependant, dès lors que l'on identifie l'objet comme étant un cœur humain, on peut expliquer son existence en faisant référence à l'histoire causale de cet organe dans l'espèece humaine ou chez les vertébrés en général. On peut ainsi rendre compte de l'existence d'un objet particulier en faisant abstraction des détails causaux le concernant par l'intermédiaire de la catégorie générale à laquelle il appartient. De la même manière, on peut expliquer le fonctionnement du système circulatoire de l'individu de la Fig. 18 (bien que ce système ne fonctionne peut-être plus au moment de prendre la photo) par l'intermédiaire du type de système correspondant chez les individus de la même espèce. Identifier une masse de tissus comme étant un cœur c'est la classer dans une catégorie fonctionnelle ; et la fonction de ce type d'organe est définie par son rôle causal dans le système circulatoire de ce type d'individus.

132 Nous avons vu au CHAP. III que cette double explication a été défendue par Griffiths (1993), Godfrey-Smith (1994) et Walsh (1996). Mais tandis que ces auteurs s'appuient sur le concept de *fitness* dans le cadre de la théorie darwinienne, notre proposition se veut plus abstraite de manière à inclure également les fonctions techniques.

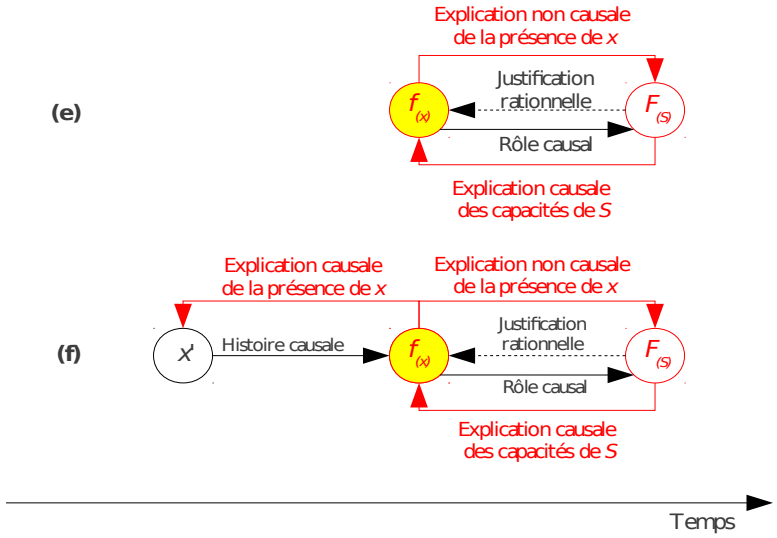

Figure 19: Schémas explicatifs de l'approche que nous proposons.

La différence et la complémentarité entre les niveaux d'explication apparaît plus clairement lorsque l'un d'eux atteint ses limites. C'est notamment le cas lorsque le porteur de fonction présente une anomalie. Si le cœur de la Fig. 18 présentait une hypertrophie, l'explication de cette particularité ne pourrait pas se limiter à des considérations générales. Une hypertrophie cardiaque peut être congénitale ou acquise[133]. Dans le premier cas, l'explication doit faire référence au patrimoine génétique des ancêtres de cet individu en particulier (ses parents, ses grands-parents, etc.) et pas aux vertébrés ni à l'espèce humaine en général. Dans le deuxième cas, l'histoire causale ne remonte pas au delà de la naissance de l'individu lui-même. On sait par exemple que l'hypertrophie cardiaque chez l'humain est souvent provoquée par une hypertension artérielle systémique, laquelle peut à son tour être causée par de nombreux facteurs génétiques, environnementaux ou comportementaux : régime alimentaire, tabagisme, résistance à l'insuline, altération de la sécrétion hormonale, etc. (Harrison, 2006) Pour expliquer l'hypertrophie hypothétique du cœur de la Fig. 18, il faudrait donc en retracer l'histoire causale particulière ou en identifier la cause spécifique. Or, des explications de cette précision étant souvent hors de portée, on se contente habituellement d'explications plus génériques exprimées en termes de causes probables, de facteurs de risque ou d'appartenance à des groupes de population.

133 Les causes les plus courantes de l'hypertrophie du ventricule gauche sont les suivantes : hypertension artérielle, myocardiopathie hypertrophique, insuffisance aortique, sténose aortique, insuffisance mitrale, sténose mitrale.

Attribuer une fonction à un trait biologique ou à une partie d'un système technique ne consiste pas seulement à en expliquer la présence, puisque cette explication est toujours possible indépendamment du caractère fonctionnel ou non de la chose expliquée, mais aussi à justifier rationnellement cette présence. Nous avons insisté à plusieurs reprises sur l'idée que la fonction d'un item est sa raison d'être (voir notamment p. 54 et suivantes). Et nous avons souligné par ailleurs l'importance que Larry Wright accorde dans sa conception des fonctions à l'idée de sélection entendue à la manière d'un choix motivé (p. 60 et suivantes). Ces deux idées sont étroitement liées. L'ingénierie inverse d'un système technique conduit à s'interroger sur les motivations de la présence des éléments qui le composent ; or ce qui motive la présence d'un élément (ce pourquoi il existe) n'est autre que sa raison d'être. L'ingénierie inverse s'applique aussi aux systèmes biologiques dans la mesure où les produits de la sélection naturelle sont à première vue comparables — et ont été longtemps confondus — avec ceux de la création artificielle. Dans les deux cas, *ce pourquoi* un item a été sélectionné est sa raison d'être, c'est-à-dire sa fonction ; mais nous avons vu avec Philip Kitcher que la réciproque n'est pas toujours vraie.

3. À propos des types fonctionnels

L'idée d'une classification fonctionnelle des traits biologiques a été critiquée notamment par Griffiths (1994, 2006) et McLaughlin (2001), défendue par Neander (2002) et discutée depuis une autre perspective par Allen (2002) et Nanay (2010). Cette polémique ne nous concerne pas, car elle porte sur la différence entre traits analogues et traits homologues, sur le rôle de la sélection naturelle et sur les méthodes classificatoires. Peu importe pour notre argument que la catégorie biologique des « cœurs » désigne tous les traits partageant la même fonction ou seulement ceux ayant un ancêtre commun mais pas forcément la même fonction. Notre argument concerne les objets (biologiques ou pas) ayant la même fonction — indépendamment de leur histoire causale et des débats sur la classification en biologie.

En effet, on peut typer des traits fonctionnels sans nécessairement avoir recours à la sélection naturelle. Nous avons vu que Buller (1998) distingue les théories étiologiques fortes pour qui les fonctions dérivent de la sélection naturelle, comme celle défendue par Neander (2002, 2010), et les théories faibles qui s'appuient sur d'autres processus non nécessairement sélectifs, comme celles défendues par Buller lui-même ainsi que par Kitcher (1993) et par Walsh (1996). Ces derniers n'adressent pas directement la question de savoir comment les traits sont typés, mais on peut déduire de leurs définitions une réponse relative à l'environnement du trait et de l'organisme porteur. Bien qu'ils parlent à

ce propos de « pressions sélectives » et de « régime sélectif », ces auteurs insistent sur le fait que la sélection naturelle n'est pas toujours suffisante ni même nécessaire. Pour Kitcher (1993, p. 271), on peut identifier la fonction d'un trait à partir de l'analyse des contraintes environnementales auxquelles il répond sans connaître pour autant les mécanismes causaux — sélectifs ou pas — responsables de son apparition. Pour Walsh (2002, p. 332), les traits fonctionnels sont des adaptations (à un environnement donné) et les adaptations peuvent apparaitre spontanément comme conséquences de la dynamique de systèmes complexes auto-organisés — avec ou sans sélection. Wimsatt (1972, 2002) examine quant à lui avec beaucoup de détail les différentes formes de similarité fonctionnelle permettant, par exemple, de comparer des traits analogues ou homologues de différentes espèces, mais sa conception, comme celle de Wright, implique un processus sélectif qui n'est pas forcément naturel[134].

Chez Wimsatt, comme chez Walsh, Kitcher, Nagel, Boorse, et d'autres, les fonctions sont définies en termes de contributions aux buts ou aux fins des organismes qui les portent. C'est-à-dire que pour attribuer une fonction à un trait, il faut d'abord identifier les fins auxquelles il contribue. Des fins comme la survie et la reproduction. Et le type auquel appartient un trait dérive notamment de ce qu'ils fait et pas — ou pas seulement — de son histoire. On rejoint de cette manière la posture que défendent certains partisans de l'approche systémique–dispositionnelle comme Joëlle Proust :

> Les convergences morphologiques que l'on observe entre des espèces sans ancêtre commun – comme celle qui existe entre l'œil du céphalopode et celui du vertébré – paraissent être partiellement constitutives de l'idée qu'une certaine fonction, définie relationnellement relativement aux fins d'un organisme complet, est l'homologue ou l'analogue d'une autre fonction, et donc détermine une classe d'équivalence appelée "fonction large". (Proust, 1995, p. 97)

Des structures différentes et non apparentées peuvent ici être classées dans un même type, à savoir un œil, parce qu'elles réalisent des contributions analogues aux mêmes fins de l'organisme, et parce qu'elles ont des caractéristiques morphologiques communes :

> [C]e qui permet de dire qu'une certaine structure de céphalopode ou de vertébré est un œil est a) qu'il permet à l'organisme de percevoir les signaux lumineux l'informant sur les propriétés de son environnement pertinentes pour son alimentation, etc. *i.e.* pour sa survie/reproduction, et b) qu'il s'agit d'une chambre noire munie

134 « Functional systems are kinds of machines, whose articulated parts contribute to the ends specified by a selection process—whether internal or external and whether of natural or artificial origin. » (Wimsatt, 2002, p. 199)

> d'une lentille, ou plus exactement, qu'il s'agit d'un dispositif de focalisation de la lumière et d'enregistrement du signal lumineux (Proust, 1995, p. 98)

On peut appliquer le même raisonnement aux artefacts. De fait, c'est ce que nous faisons habituellement face à un artefact nouveau. On déduit la fonction des éléments qui le composent à partir de leur similarité avec d'autres items connus et de leur contribution à la finalité présumée de l'artefact. C'est aussi ce que font, plus rigoureusement, les archéologues qui étudient des objets comme le mécanisme d'Anticythère.

L'ingénierie inverse, défendue par Dennett (1990), permet de classer les artefacts et leurs parties dans des catégories fonctionnelles en s'intéressant à leur organisation et aux intentions présumées de leurs créateurs plutôt qu'à leur histoire causale. Et la substitution de la référence à l'intentionnalité par une référence à la finalité permettrait d'appliquer au domaine biologique le raisonnement par rétro-ingénierie.

4. Conséquences fonctionnelles et conséquences accidentelles

Nous avons vu au CHAP. I que la distinction fonction/accident est fondamentale pour les partisans de la conception étiologique et que le recours à la sélection naturelle ou, plus généralement, à une approche diachronique permet de pallier à ce défaut en écartant les situations où la sélection résulte d'un processus accidentel. Mais l'approche diachronique n'est pas forcément la seule ni la meilleure manière d'établir cette distinction. En analysant l'organisation d'un système technique, on peut dans la plupart des cas déterminer assez facilement si la présence des éléments qui le composent est accidentelle ou pas, indépendamment de l'histoire causale du système et indépendamment du concept de fonction. De la même manière, on peut déterminer le caractère accidentel ou pas des conséquences de la présence d'un item par le biais d'une analyse synchronique du système auquel il appartient[135]. C'est ce que fait l'ingénierie inverse lorsqu'elle cherche à déterminer la raison d'être des éléments qui

135 Dans l'exemple utilisé par Wright de la boucle de ceinturon qui arrête les balles, cette conséquence est clairement accidentelle puisque rien ne la laisse prévoir : elle offre une surface protectrice très réduite et située contre toute logique là à un endroit où elle ne protège aucun organe vital, de manière que l'impact d'une balle à cet endroit ne peut être que fortuit. Autrement dit, quand on analyse cet item à la lumière de cette finalité (offrir une protection contre les balles), le moins que l'on puisse dire est que son implémentation est suboptimale. Par conséquent, à défaut d'indices d'un usage intentionnel dans ce sens, il n'y a aucune raison de penser que les boucles de ceinturon aient été conçues ou utilisées pour dévier les balles.

le composent. c'est-à-dire qu'il est possible d'établir la distinction fonctionnel–accidentel requise par l'approche étiologique sans adopter la perspective diachronique défendue par ses partisans. Mais il convient à nouveau de distinguer entre le système fonctionnel abstrait et les éléments matériels concrets qui l'implémentent. Dans l'exemple donné par Kitcher (1993, p. 260) de la petite vis tombée par inadvertance dans une machine en construction, la présence de la vis est accidentelle, mais la présence d'une jonction à cet endroit est nécessaire pour que la machine fonctionne. En tant que jonction, c'est-à-dire en tant qu'élément fonctionnel $f_{(X)}$, la vis a donc une raison d'être dans le système et on peut justifier rationnellement sa présence bien que, par ailleurs, en tant qu'objet matériel concret (*token* de type *X*), sa présence soit accidentelle[136].

En ce qui nous concerne, cela veut dire que même si l'hypertrophie ventriculaire n'avait pas été historiquement sélectionnée pour répondre à une hypertension chronique et si, par conséquent, sa présence dans ces conditions était en quelque sorte accidentelle, elle pourrait néanmoins être justifiée par une analyse fonctionnelle du système circulatoire. Une telle analyse consiste à examiner l'hypertrophie ventriculaire de la même manière que la vis de l'exemple précédent, c'est-à-dire à déterminer si sa présence et les conséquences de sa présence font partie intégrante du système ou si, au contraire, elles relèvent d'un concours de circonstances. Cette analyse ne se limite pas, comme chez Cummins, à étudier le rôle causal que joue un item relativement à une capacité du système arbitrairement choisie par l'observateur ; elle consiste également, comme dans l'exemple du mécanisme d'Anticythère, à identifier la capacité du système pour laquelle un item a été conçu. De même que l'organisation d'un artefact est révélatrice des intentions ayant guidé sa construction, l'organisation d'un être vivant est révélatrice des pressions sélectives que ses ancêtres ont subi.

Pour illustrer sa distinction entre conséquences fonctionnelles et conséquences accidentelles, Larry Wright (1973, p. 147) donne l'exemple de la trotteuse d'une montre qui, bien qu'elle puisse servir — accidentellement — à retirer la poussière des chiffres du cadran, n'a évidemment pas cette fonction. Mais imaginons un pays *P* tellement poussiéreux que les horlogers y fabriquent des montres dont la trotteuse a volontairement les deux fonctions, celle d'indiquer les secondes et celle de retirer la poussière. Pourquoi les trotteuses les montres de *P* ont-elles les deux fonctions et pas celles fabriquées ailleurs ? On pourrait répondre que la différence se situe dans l'intention des horlogers. Mais cette réponse n'est pas tout à fait satisfaisante. Si les trotteuses des montres de *P* ont pour fonction de

136 En termes aristotéliciens nous distinguerions la cause finale–efficiente, à savoir l'organisation fonctionnelle de la machine, et la cause matérielle–efficiente de la présence de la vis, qui est ici accidentelle.

retirer la poussière des chiffres du cadran, alors leur *design* est vraisemblablement optimisé pour cela. Par conséquent, un horloger d'un pays tiers qui examinerait pour la première fois les montres de *P* sans en connaître la provenance ne manquerait probablement pas de remarquer la singularité de leur conception et s'interrogerait sans doute sur les motivations de ses confrères. Par la seule analyse du système, il pourrait éventuellement découvrir la fonction seconde des trotteuses de ces montres, celle de retirer la poussière, car c'est ce pour quoi elles ont été optimisées. C'est-à-dire que s'il raisonnait, par exemple, en termes de problèmes et solutions, il pourrait découvrir le problème pour lequel les trotteuses de ces montres constituent la meilleure solution[137]. De cette manière, il pourrait aussi en inférer les intentions de leurs créateurs ; mais il convient de signaler que la vérité de cette seconde inférence ne conditionne pas nécessairement la validité de l'attribution fonctionnelle. En effet, si les objets en question n'étaient pas le produit d'un dessein mais celui de la sélection naturelle (il est évidemment plus facile d'imaginer cela pour d'autres objets que des montres), alors le raisonnement en termes de problèmes et solutions continuerait à être valable. Nous voulons montrer par là que la différence entre conséquences fonctionnelles et accidentelles repose — au moins en partie — sur le *design* du système lui-même, c'est-à-dire sur le fait que celui-ci semble avoir été conçu pour obtenir ces conséquences. Tandis que les trotteuses des montres de *P* semblent avoir été conçues pour retirer la poussière des chiffres du cadran, les boucles de ceinturon, au contraire, ne semblent pas avoir été conçues pour dévier ni pour arrêter les balles de revolver. Et bien que toutes les trotteuses puissent éventuellement servir à retirer la poussière, seules celles de *P* sont optimisées pour cela, indépendamment du mécanisme causal responsable de cette optimisation (création intentionnelle ou sélection naturelle).

5. Le problème de l'adaptationnisme et l'optimalité

Comme nous l'avancions au CHAP. III, SECT. 4, notre analyse repose donc en partie sur une recherche d'optimalité, laquelle est sujette à controverses en biologie depuis la publication de Gould et Lewontin

137 Nous ne voulons pas dire que cette solution-là soit la meilleure possible pour résoudre le problème de la poussière (il est suffisant que ce soit une bonne solution), mais que le problème de la poussière est celui pour lequel les trotteuses de ces montres offrent la meilleure solution. Autrement dit, si l'on envisage une série de possibles problèmes auxquels les trotteuses de ces montres pourraient apporter la solution, le problème de la poussière est sans doute celui pour lequel la solution apportée est la meilleure..

(1979) critiquant l'adaptationnisme[138]. Or, nous nous en écartons sur un point essentiel, à savoir, sur le rôle de la sélection naturelle dans le processus évolutif. En effet, tandis que l'adaptationnisme met l'accent sur la sélection naturelle comme moteur principal — si ce n'est exclusif — de l'évolution et de l'optimisation des traits fonctionnels (Orzack & Forber, 2010), nous cherchons au contraire à faire abstraction des mécanismes causaux qui en sont responsables[139]. De cette manière, nous nous situons en dehors du débat sur l'adaptationnisme proprement dit tout en restant concernés par les critiques portant sur ce que Tim Lewens (2002) appelle le modèle de l'artefact en biologie.

En réponse à l'une de ces critiques, Parker & Maynard Smith (1990) avaient déjà essayé de montrer que l'adaptationnisme ne postule pas ni ne requiert une adaptation optimale des êtres vivants ou de leurs parties. Nonobstant, le critère d'optimalité se révèle souvent utile pour répondre à certaines questions liées au travail d'optimisation de la sélection naturelle. Il peut permettre de comprendre, par exemple, pourquoi le ratio mâles/femelles dans une population est souvent proche de

138 D'après Rose et Lauder (1996), les critiques portant sur l'adaptationnisme ont commencé dans les années 1960 et s'appuyaient sur des raisons aussi bien conceptuelles qu'empiriques. Pour une analyse rétrospective de l'article de Gould et Lewontin ainsi que ses retombées au cours des deux décennies suivantes, voir Pigliucci et Kaplan (2000).

139 L'une des principales critiques de Gould et Lewontin concernait l'idée selon laquelle il existe une explication adaptative pour chaque trait. Cette idée comporte elle-même plusieurs présupposés parmi lesquels nous voulons dégager les trois suivants. Le premier est celui de l'utilité présente ou passé des traits biologiques. Le second est leur optimalité. Le troisième est leur origine sélective. D'après les auteurs cités, les travers de l'adaptationnisme se situent dans les deux derniers présupposés et en particulier dans leur association. Ils ne discutent pas, par exemple, le fait que les pattes antérieures des Tyrannosaures aient pu leur être utiles mais plutôt l'explication de leur petitesse comme étant forcément la solution optimale trouvée par la sélection naturelle à un problème adaptatif. Dans la mesure où nous faisons abstraction des causes — aussi bien proximales que distales — responsables de la taille des pattes des Tyrannosaures, la critique de Gould et Lewontin ne nous concerne pas. Godfrey-Smith (2001) distingue de son côté trois formes d'adaptationnisme : empirique, explicatif et méthodologique. La première affirme que la sélection naturelle est le moteur causal principal de l'évolution biologique ; la seconde affirme que l'explication du *design* apparent des organismes est le problème principal de la biologie et que la sélection naturelle en est la réponse ; la troisième se contente d'affirmer que la meilleure façon de comprendre les systèmes biologiques consiste à les envisager en termes d'adaptation et de *design*. Cette dernière est neutre quant à la sélection naturelle et ne considère l'adaptation et le *design* que d'un point de vue heuristique. C'est la seule qui corresponde à l'approche que nous voulons développer ici.

1:1 ou au contraire pourquoi il s'écarte de cette valeur dans une population donnée. Cela ne veut pas dire que la sélection naturelle soit toujours adaptative (Rose & Lauder, 1996) ni que tous les phénomènes biologiques aient forcément une explication en ces termes, ni que les phénomènes biologiques soient optimaux ou seulement optimisés (Maynard Smith, 2006), mais que, lorsqu'un phénomène est effectivement adaptatif, le critère d'optimalité peut s'avérer utile — et constitue parfois la seule option raisonnable — pour trouver l'explication correspondante (Orzack & Sober, 1994; P. Abrams, 2001, p. 274). De la même manière, il serait absurde de prétendre que tous les éléments d'un système technique sont des réponses optimales à des problèmes donnés. En effet, outre les éléments aléatoires et les contraintes de type *spandrel*, il est probable que deux ingénieurs différents trouveront des solutions différentes pour un même problème technique, et il est vraisemblable qu'aucune de ces solutions ne soit optimale bien que, par ailleurs, chacune d'elles soit la meilleure solution qu'ils aient trouvé dans la limite de leurs compétences et des connaissances techniques propres à leur culture, du temps imparti et des ressources disponibles.

Il n'en demeure pas moins que certaines solutions sont meilleures que d'autres et que l'optimalité apparaît dès lors comme un bon critère par défaut pour identifier les solutions correspondant à un problème donné (pensée adaptative) ou les problèmes correspondant à une solution (ingénierie inverse)[140]. C'est-à-dire que parmi toutes les solutions possibles pour un même problème (et vice-versa), celle que l'on cherche fait probablement partie des meilleures disponibles.

Considérons par exemple un item x dont on cherche à connaître la fonction. Si cet item est un artefact ou une partie d'un artefact, alors parmi ses usages possibles α, β, γ, δ, celui pour lequel il a été créé ou pour lequel il est employé correspond probablement à son usage optimal (Fig. 20). Une tasse à café (x) peut ainsi servir comme projectile (α), comme marteau (γ) ou comme presse-papiers (δ), parmi d'autres usages possibles, mais elle s'avère plus utile comme récipient pour porter à la bouche un liquide chaud (β), car c'est ce qu'on peut faire de mieux avec[141] et parce que, parmi d'autres objets w, y, z, c'est celui qui s'acquitte

140 Ce qui ne veut pas dire, nous insistons sur ce point, que tous les traits biologiques soient forcément des solutions à des problèmes adaptatifs. Gould et Lewontin (1979) critiquaient justement la tendance de l'adaptationnisme à se raconter des histoires, c'est-à-dire à toujours imaginer une histoire adaptative pour chacun des traits d'un organisme et à remplacer une histoire défaillante par une autre du même type sans qu'il soit possible de la vérifier.

141 En tant que projectile, la tasse est un objet relativement peu efficace et peu efficient, car elle n'est pas particulièrement ergonomique ni aérodynamique, et elle est inutilement compliquée. En tant que presse-papiers, sa complexité n'a pas non plus de justification rationnelle, car n'importe quel objet suffisamment lourd à bord plat pourrait servir aussi bien. Et en tant que mar-

le mieux de cette tâche[142]. C'est ce type de raisonnement qui peut permettre d'identifier la fonction seconde des trotteuses des montres du pays *P* ou, plus sérieusement, la fonction des engrenages du mécanisme d'Anticythère. Ce type de raisonnement est également applicable aux traits biologiques (l'œil, la chlorophylle, etc.) abstraction faite des mécanismes causaux sous-jacents.

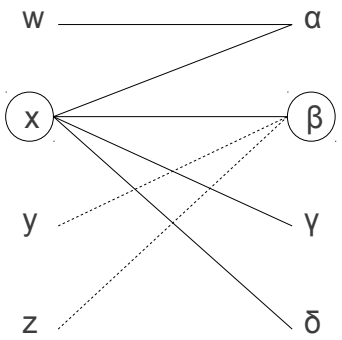

Figure 20: Correspondances multiples entre un item et ses usages.

Une autre manière de concevoir la question de l'optimalité consiste à penser en termes de balance coûts-bénéfices. La présence d'un item implique certains coûts (de fabrication, de fonctionnement, de maintien) et apporte certains bénéfices (ce en quoi il est utile). Or, s'il est rationnel de chercher à maximiser les bénéfices au moindre coût dans le domaine des activités humaines, il semble également raisonnable de s'attendre à observer un phénomène analogue dans le domaine naturel. Par exemple, pourquoi le patrimoine génétique des êtres vivants terrestres repose-t-il sur des molécules aussi complexes que l'ARN et l'ADN alors qu'il existe des molécules informationnelles plus simples pouvant jouer le même rôle (Orgel, 1998, 2004) ? On pourrait s'attendre à ce que la Nature favorise la solution la plus simple et la moins coûteuse pour résoudre un problème donné, et d'après certains spécialistes, la Nature aurait effectivement employé ces molécules à un stade précoce de l'apparition de la vie, avant l'émergence d'un monde à ARN et avant que ce dernier ne donne naissance aux organismes à ADN, ce qui laisse à penser que les molécules les plus récentes seraient d'une manière ou d'une autre plus avantageuses que les précédentes[143]. Et s'il existait d'autres molécules informationnelles

---

teau, rien ou presque ne plaide en sa faveur. En revanche, sa forme et sa composition sont parfaitement adaptées pour l'usage qu'on lui donne, à savoir, pour boire un café, un thé ou n'importe quel autre liquide chaud : la céramique est imperméable et isolante, la anse permet de la tenir sans se brûler et le léger rebord de la partie inférieure minimise la surface de contact avec la table (pour ne pas la brûler ?).

142 En tant que projectile (α) la tasse (x) est moins efficace et moins efficiente que d'autres objets plus simples (w) tels que cailloux, balles, etc. En revanche, en tant que récipient pour le thé ou le café, elle est plus adaptée que la plupart des objets que l'on peut imaginer. C'est sans doute l'une des raisons pour lesquelles nous avons tendance à employer des tasses pour boire le café plutôt que des cailloux, des assiettes, des fourchettes, etc.

143 Il n'est pas facile de préciser par rapport à quoi elles sont plus avantageuses, si ce n'est de façon générale en termes de *fitness*, mais cela n'a pas d'impor-

encore plus « avantageuses » que l'ADN[144] qui ne se situent pas dans la même lignée évolutive, on pourrait justifier leur non-utilisation en alléguant ou bien que la nature n'a pas eu le temps de les découvrir, ou bien qu'il serait plus coûteux d'effectuer une transition vers ces nouvelles molécules que de conserver les anciennes.

Le critère d'optimalité est compatible avec l'idée de François Jacob (1977) selon laquelle la Nature agit non pas comme un ingénieur mais comme un bricoleur qui, au lieu de concevoir et construire *ex novo*, se contente d'adapter l'existant aux nouveaux usages qu'il veut lui donner (voir CHAP. III, SECT. 4.2). De cette manière, bien que le résultat du processus dans son ensemble soit suboptimal comparé à l'hypothétique design d'un ingénieur idéal, il n'en demeure pas moins la meilleure ou l'une des meilleures solutions envisageables à chaque étape du processus à partir de la situation immédiatement antérieure. Suivant la théorie endosymbiotique de Lynn Margulis, la principale source d'innovation évolutive serait d'ailleurs non pas les mutations aléatoires mais l'incorporation symbiotique d'éléments précédemment perfectionnés dans des lignées séparées (Margulis, 1981; Margulis & Sagan, 1995). Et tout nouvel organisme étant nécessairement issu d'un organisme précédent dont il hérite l'essentiel des traits, il est évidemment impossible de remettre à plat l'organisation d'un système pour le reconstruire à partir d'une solution idéale. Il n'est pas étonnant, dans ces conditions, que d'anciens traits assument de nouvelles fonctions pour lesquelles ils n'ont pas été conçus et que d'autres, bien qu'ayant perdu tout ou partie de leur fonctionnalité, continuent à être présents dans le système.

L'application rigoureuse du critère d'optimalité lors de l'analyse d'un système biologique ou technique semble donc requérir la prise en considération de sa dimension historique et ne peut pas s'appuyer uniquement sur ce que Griffiths (1996) appelle des généralisations adaptatives ou généralisations fonctionnelles[145]. Mais, dans le cadre d'une analyse reposant sur le modèle de l'artefact, la connaissance de la réalité historique

---

tance particulière pour notre argument.

144 Par exemple en termes de stabilité et de réplication (Kool, Morales, & Guckian, 2000).

145 « The adaptationist supposes that there are adaptive (or 'functional' or 'ecological') generalizations that can explain the existence of certain biological forms. These generalizations rank alternative traits in terms of their fitness. The actual trait is explained by citing generalizations that assign it a higher fitness than the alternatives. […] Successful adaptationist explanations would require adaptive generalizations that are insensitive to historical and to the complete range of characters present, and robust under stochasticity. Only this would make it possible to explain the actual character which results by citing the functional generalizations alone. » (Griffiths, 1996, p. 515).

n'est pas essentielle pour comprendre le fonctionnement du système[146]. Il est vrai, comme dit Griffiths, que les généralisations fonctionnelles donnent lieu à une prolifération d'hypothèses alternatives sous-déterminées par les faits sur lesquels elles s'appuient, un même trait pouvant recevoir plusieurs explications adaptatives différentes également compatibles avec les contraintes écologiques supposées responsables de son apparition. Et il est vrai aussi que la confrontation de ces hypothèses avec les faits historiques peut servir, comme l'indique cet auteur, pour en limiter la prolifération. Mais la multiplicité des hypothèses ne constitue un réel problème que dans la mesure où l'on s'intéresse aux causes de la présence d'un trait, car si l'on s'intéresse plutôt à ses raisons on conviendra qu'un même trait peut parfaitement avoir plusieurs raisons d'être. De la même manière, en littérature, on peut s'intéresser à ce que l'auteur a voulu dire dans un texte donné, ce qui implique une analyse minutieuse non seulement du texte lui-même mais aussi de la biographie de l'auteur et des circonstances historiques précises au moment de sa rédaction. Nonobstant, on peut aussi s'interroger sur les autres significations possibles du texte et sur les raisons d'être des éléments qui le composent indépendamment des intentions originales de son auteur. Il s'agit là de deux démarches différentes répondant à des intérêts différents bien que complémentaires. L'analyse des significations secondes peut révéler certaines choses que l'auteur ne souhaitait pas voir divulguées ou dont il n'était pas conscient, et une bonne connaissance de l'auteur peut réciproquement éclairer certains aspects de son œuvre autrement incompréhensibles car liés au hasard ou à des circonstances particulières[147]. Dans le domaine biologique, certains traits n'ont pas de raison d'être au-delà du hasard ou de certaines circonstances particulières que seule une analyse historique peut éventuellement dévoiler, mais la

146 Au contraire, la compréhension préalable du fonctionnement des parties d'un organisme (ou d'un artefact) est la meilleure — et parfois la seule — manière de comprendre son étiologie, comme l'indique Michael Ruse : « *Rather than being thrown back at once on impossible questions about how natural selection operated in the distant past, evolutionists begin by asking how things work at the moment. When once they have got a handle on this question, they are able next to talk in terms of natural selection, trying to relate their studies back to what happened in the past. But this second stage can occur only if first they have used their design metaphor to ask pertinent questions about function.* » (Ruse, 2002, p. 47).

147 À titre d'anecdote, Quentin Tarantino raconte que les critiques de cinéma ont beaucoup spéculé à propos du ballon de couleur qui apparaît dans la dernière scène de *Reservoir Dogs*, lui cherchant une signification cachée en relation avec les codes de couleur que se donnent les personnages dans le film, alors qu'en réalité le ballon s'est tout simplement échappé accidentellement des mains d'un enfant qui assistait parmi le public au tournage de cette scène.

connaissance historique est ici considérée comme un outil au service de la compréhension plus générale du fonctionnement du système plutôt que comme une fin en soi, au même titre que le critère d'optimalité.

Outre la multiplicité des hypothèses alternatives, un autre aspect de la critique de Griffiths est que la pensée adaptative et l'ingénierie inverse conduisent facilement à des conclusions erronées lorsque l'historicité et la spécificité des êtres vivants ne sont pas prises en considération. Il cite en particulier l'argument de la convergence évolutive employé par Dennett (1995) pour justifier certaines généralisations fonctionnelles indépendantes des circonstances historiques. Selon Dennett, si les yeux de la plupart des animaux ont un dispositif sensible aux formes manifestant une symétrie autour d'un axe vertical, alors ce dispositif doit avoir une utilité très générale correspondant à un facteur environnemental commun et pouvant servir par exemple pour détecter rapidement la présence d'un autre organisme regardant de face, puisque la plupart des animaux possèdent en effet une symétrie verticale. Or, selon Griffiths (1996, p. 522), il pourrait s'agir au contraire d'une homologie chez les vertébrés, indiquant que cette caractéristique est apparue une seule fois dans l'histoire et a ensuite été transmise à tous les descendants. Elle pourrait alors remplir des fonctions différentes dans différents groupes, voir même être présente par inertie phylogénétique sans aucune fonction particulière. Mais la critique de Griffiths est à la fois juste et assez peu convaincante, car l'erreur de Dennett, si erreur il y a, concerne d'une part l'origine du trait (convergence évolutive vs. descendance commune) et d'autre part son utilité actuelle (générale vs. spécifique), or aucune de ces deux erreurs ne dérive directement de l'application de l'ingénierie inverse et, bien que la première puisse être évitée grâce aux méthodes historico-comparatives défendues par Griffiths, la seconde requiert seulement une analyse minutieuse de chacun des systèmes. Griffiths a certes raison de mettre en garde contre les errements auxquels peuvent mener la pensée adaptative et l'ingénierie inverse appliquée aux êtres vivants, mais comme le souligne très justement Tim Lewens (2002, 2004), les mêmes problèmes apparaissent dans l'analyse des artefacts.

Si le modèle de l'artefact est tout aussi problématique avec les artefacts qu'avec les êtres vivants, alors les difficultés inhérentes à son utilisation ne sont pas invalidantes. Au contraire, puisque la pensée adaptative et l'ingénierie inverse comptent, malgré ces difficultés, parmi les meilleures techniques d'analyse fonctionnelle appliquées aux artefacts, rien ne s'oppose à ce qu'elles soient également efficaces pour les êtres vivants. Dans un cas comme dans l'autre, la présupposition d'optimalité semble être un bon critère d'attribution fonctionnelle par défaut (Dennett, 1995, p. 212-213) qui ne dépend pas, comme nous avons essayé de le montrer, ni de l'origine ni de la nature des mécanismes causaux responsables de la présence du trait ou du système en question.

Chapitre VII

Définition téléologique du concept de fonction

L'exercice d'une activité fonctionnelle, qu'elle soit naturelle ou artificielle, est directement sensible au contexte. Ruth G. Millikan souligne très justement qu'une analyse fonctionnelle à la Cummins appliquée à un système idéal, c'est-à-dire à un type de système, ne peut pas ne pas tenir compte de ses conditions extérieures de fonctionnement :

> « The circuit diagram that comes with your washing machine represents how it was designed or intended to function, not necessarily how it does function. Moreover, it was designed to function that way not inconditionally, but given quite specific background conditions and quite specific inputs. For example, it was designed to operate upright on a relatively level, stable, and rigid floor, under about one g gravitational force, surrounded by air at about one atmosphere pressure, protected from large magnetic forces, heavy blows, strong vibrations, heavily corrosive gases, and so forth. […] Specifications of this sort concerning background conditions of operation and allowable input must also be assumed for any system to be given a Cummins-style functional analysis—we can say, for any 'Cummins system'. » (Millikan, 2002, p. 119-120)

Dans une conception darwinienne, la sélection d'un trait est relative à l'environnement dans lequel vivent les porteurs de ce trait. Cependant, lorsque la définition est tournée vers le passé, le fait avéré de la sélection suffit à justifier l'attribution fonctionnelle de sorte que la référence au contexte environnemental n'est pas nécessaire. En revanche, les partisans d'une approche propensionniste lui accordent une place importante dans leurs définitions. Nous allons donc nous appuyer sur ce type d'approche pour proposer une définition sensible au contexte qui soit applicable à la fois aux êtres vivants et aux artefacts.

Notre point de départ sera la définition relationnelle de Walsh déjà commentée au CHAP. I, car cette définition vise non seulement à réunir dans une formulation unique les conceptions étiologique et propensionniste, voire dispositionnelle, rendant possible une double explication de la présence d'un trait et de la *fitness* des organismes porteurs, mais elle est également celle qui s'appuie le plus ouvertement sur le contexte environnemental entendu en termes de régime sélectif (ce qui la rapproche par ailleurs de la proposition de Kitcher).

1. Définition téléologique des fonctions

La définition originale de Walsh est la suivante :

> « The/a function of a token of type *X* with respect to selective regime *R* is to *m* iff *X*'s doing *m* positively (and significantly) contributes to the average fitness of individuals possessing *X* with respect to *R*. » (Walsh, 1996, p. 564)

Pour pouvoir l'appliquer aux artefacts, nous proposons de remplacer la notion de régime sélectif par la notion plus générale de contexte, la notion de *fitness* individuelle moyenne par celle de fin, et la notion d'individu par celle de système :

> La/une fonction d'un *token* de type *X* dans le contexte *C* est *f* si la réalisation de *f* par *X* contribue à une fin *F* du système *S* auquel *X* appartient.

Mais nous avons insisté plus haut sur l'idée qu'un système fonctionnel *S* n'est pas constitué par des structures et des processus matériels *X* mais par les fonctions qui leur correspondent. Cela nous conduit à la formulation suivante, où la notation $f_{(X)}$ signifie *fonction de X* :

> La/une fonction d'un *token* de type *X* dans le contexte *C* est $f_{(X)}$ si la réalisation de $f_{(X)}$ par *X* contribue à une fin *F* du système *S* auquel $f_{(X)}$ appartient.

Pour être plus précis et plus rigoureux, nous devrions préciser, suivant Wimsatt (1972, 2002), que les fonctions ne sont pas attribuables aux structures elles-mêmes mais à leurs comportements, opérations ou processus. Ainsi, nous ne devrions pas attribuer de fonction au cœur, en tant que structure matérielle, mais ses battements. L'un des avantages de cette proposition, de notre point de vue, est qu'elle évite de faire des fonctions des propriétés intrinsèques d'objets physiques (Gayon, 2010b).

Par ailleurs, puisque l'on attribue une fonction à l'activité d'un *token* en vertu du type auquel il appartient, il nous semble également préférable de l'attribuer relativement à un type de système plutôt qu'à un

système particulier, bien que celui-ci puisse être unique en son genre. Cela donne la définition corrigée suivante :

> La/une fonction (de l'activité) d'un *token* de type *X* dans un système de type *S* et dans le contexte *C* est $f_{(X)}$ si la réalisation de $f_{(X)}$ par *X* contribue à une fin *F* du système *S* auquel $f_{(X)}$ appartient.

On peut l'améliorer en distinguant les trois conditions qui la composent. La première est la réalisation de $f_{(X)}$ par *X*, entendue comme une réalisation effective ou comme une capacité. En renversant la formulation, on dira que $f_{(X)}$ est une conséquence (résulte) de l'existence de *X*. On retrouve ainsi l'une des conditions de la définition de Wright. La seconde est le fait que cette conséquence contribue à une fin du système et s'inscrit donc dans une relation de type moyens–fin. La troisième est son appartenance au système, c'est-à-dire que la contribution en question n'est pas accidentelle mais constitutive du système. Cette troisième condition inscrit la fonction dans une relation parties–tout.

DÉFINITION :

La/une fonction (de l'activité) d'un *token* de type *X* dans un système de type *S* et dans le contexte *C* est $f_{(X)}$ si :
(i) $f_{(X)}$ est une conséquence de *X* ;
(ii) $f_{(X)}$ contribue à une fin du système $F_{(S)}$;
(iii) $f_{(X)}$ appartient à *S*.

2. Compatibilité avec d'autres conceptions

Cette formulation vise à réunir et à généraliser celles de Larry Wright (1973) et de Robert Cummins (1975) et à assimiler, à travers elles, les autres conceptions qui en découlent.

On se souvient que l'analyse fonctionnelle de Cummins consiste à expliquer une capacité d'un système à travers sa décomposition en sous-systèmes et à étudier leurs contributions causales respectives dans la réalisation de cette capacité. On se souvient aussi qu'une interprétation de cette analyse en termes de *tokens* donne lieu à des conséquences contre-intuitives, puisqu'un cœur malformé ou malade, incapable de pomper le sang, se voit alors privé de fonction. La critique de cette interprétation nous avait amenés à proposer une interprétation alternative en termes de types qui nous semble conforme à l'esprit de la conception disposition-

nelle (voir CHAP. II, p. 106). Au lieu de porter sur un objet concret, nous situons l'analyse à un niveau plus abstrait depuis lequel il est possible d'attribuer une capacité au mécanisme d'Anticythère et une fonction à ses parties bien que le mécanisme en question ne soit aujourd'hui qu'un amas de fragments métalliques rouillés. Mais au lieu d'analyser un système biologique ou technique (*S*) comme étant un ensemble d'éléments matériels (*X*) dotés de fonctions (*f*) rendant possible une capacité, on peut le voir plutôt comme un ensemble de fonctions ($f_{(X)}$) implémentées matériellement (*X*). Lorsqu'on décrit le système circulatoire comme étant un circuit fermé comportant une pompe, des filtres, des échangeurs chimiques, etc. (Fig. 16, p. 235), on mentionne des éléments fonctionnels indépendants de leurs concrétions matérielles (Fig. 18, p. 236). De ce fait, si l'on substitue un cœur biologique par un cœur artificiel et que l'on pallie une insuffisance rénale avec un dialyseur externe et une circulation extracorporelle, on ne change pas le système circulatoire en tant que système fonctionnel ; ce qui change ce sont les éléments matériels qui implémentent ces fonctions.

La définition téléologique que nous proposons s'appuie donc sur une analyse fonctionnelle généralisée. Si l'on entend la fin *F* comme étant une capacité du système *S*, alors l'attribution d'une fonction $f_{(X)}$ à un *token* de type *X* consiste à dire que *S* se compose d'un ensemble de fonctions $\{f_{(a)}, f_{(b)}, \ldots, f_{(x)}, \ldots\}$ organisées de telle sorte qu'elles rendent collectivement possible la capacité $F_{(S)}$ dans le contexte *C* (voir Fig. 21), et que *X* est l'implémentation de l'une de ces fonctions, $f_{(x)}$. De cette manière, conformément à la conception de Cummins, attribuer une fonction à un item *X* revient à lui attribuer un rôle causal dans la réalisation d'une capacité du système que l'on cherche à expliquer.

On remarquera que si le choix de la fin $F_{(S)}$ est arbitraire, la définition téléologique est sujette à la même critique que celle de Cummins, à savoir, celle d'attribuer au cœur la fonction de faire du bruit dans la mesure où il contribue au « système sonore » de l'organisme. Mais là réside justement l'une des différences principales entre notre proposition et celle de Cummins : tandis que le choix de la capacité étudiée dépend de l'observateur et n'est limité que par des critères de pertinence explicative, celui de la finalité (qui peut être une capacité) est soumis à d'autres types de contraintes. On peut justifier scientifiquement l'affirmation courante selon laquelle les êtres vivants ont pour finalité la survie et la reproduction, et à partir de là on peut analyser tous les systèmes et sous-systèmes de l'organisme du point de vue de leur contribution à cette double fin. Mais il semble beaucoup plus difficile de justifier l'affirmation selon laquelle les êtres vivants auraient pour finalité de louer le Seigneur en faisant du bruit, pour reprendre l'exemple de Searle (1995, p. 15).

On remarquera aussi que notre définition préserve la distinction entre la fonction $f_{(X)}$ d'un item *X* dans un système *S* et le fait qu'il puisse *fonctionner comme Y* dans un système S'. Une distinction que

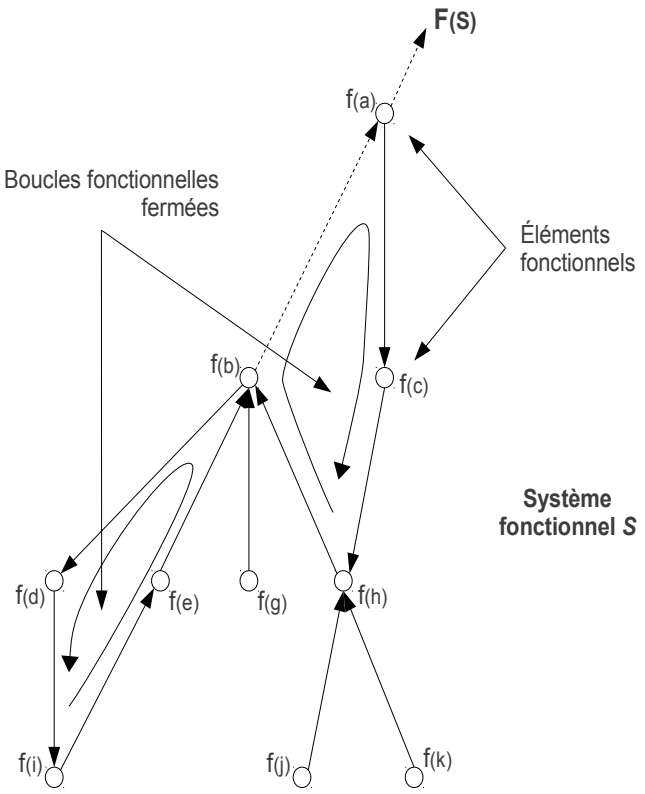

Figure 21: Schéma d'un système fonctionnel bouclé. Système fonctionnel *S* composé d'un ensemble de fonctions *f(x)* qui collectivement rendent possible la finalité *F(s)* dans un contexte *C*. (Source : adaptation à partir de Wimsatt 2002).

Boorse (1976, p. 81) traduit en termes de « fonction possédée » (*function possessed*) et de « fonction réalisée » (*function performed*). La différence entre les deux se situe, de notre point de vue, dans la finalité du système auquel l'item *X* appartient. Les battements du cœur contribuent à un hypothétique système sonore de l'organisme, mais ce système n'ayant aucune finalité rationnellement attribuable sur une base non arbitraire, notre définition ne leur est pas applicable. On pourrait donc dire qu'ils fonctionnent comme base rythmique de ce système sonore, mais on ne peut pas dire que ce soit là leur fonction, car ils ne contribuent à aucune fin. En revanche, les battements cardiaques ont bel et bien une fonction de base rythmique, conformément à notre définition, dans la chanson « Speak to me » de *Pink Floyd*.

On remarquera enfin que cette définition rend compte de la fonction de la petite vis tombée par inadvertance dans la machine de l'exemple de Kitcher. Étant donné qu'il serait facile d'imaginer des exemples biologiques analogues pour essayer de mettre à mal la définition

précédente, nous insistons à nouveau sur le fait que l'attribution fonctionnelle, ici, n'est pas et ne prétend pas être une explication causale de la présence du trait sur lequel porte l'attribution.

Cette dimension non causale de l'explication est l'une des différences principales entre notre définition et les définitions étiologiques. Cependant, nous avons vu plus haut que l'attribution fonctionnelle peut être à la fois une explication causale des capacités d'un système et une explication non causale de la présence des traits qui le composent. Nous avons vu aussi que la conception étiologique, en particulier celle de Larry Wright, pouvait être interprétée non pas seulement en termes de causes mais aussi plus généralement en termes de raisons : la fonction d'un item est sa raison d'être dans le système, et lorsque cette présence est le résultat d'une sélection, alors la fonction est ce pour quoi il a été sélectionné. Depuis cette perspective, notre définition rejoint celle de Wright que nous reformulons comme suit :

> La fonction de *X* (dans le système *S* et dans le contexte C) est $f_{(X)}$ si :
> 1. $f_{(X)}$ est une conséquence de l'existence de *X* (dans *S*) ;
> 2. *X* existe ou est présent (dans *S*) parce qu'il réalise $f_{(X)}$ (dans C).[148]

La première condition de la définition de Wright est présente dans la première condition de la définition téléologique. On retrouve la seconde à partir du raisonnement suivant. L'analyse fonctionnelle du système *S* nous apprend que celui-ci comprend parmi d'autres la fonction $f_{(X)}$; or celle-ci est implémentée matériellement par l'item *X* ; par conséquent, on peut dire que *X* est présent (dans *S*) parce qu'il réalise $f_{(X)}$ et que $f_{(X)}$ appartient à *S*. Par exemple, l'analyse du système circulatoire sanguin (*S*) révèle que celui-ci comprend une pompe ($f_{(X)}$) ; or il se trouve que cette fonction est réalisée par le cœur (*X*) ; on dira donc, conformément à la seconde condition de la définition de Wright, que le cœur (*X*) est présent dans le système (*S*) parce qu'il pompe le sang ($f_{(X)}$). Peu importe que le cœur en question soit naturel ou artificiel, peu importe qu'il soit le résultat d'un processus de sélection naturelle ou de fabrication consciente, et peu importe que le *token* de type cœur (*X*) qui remplit la fonction ($f_{(X)}$) soit là par hasard (comme la petite vis de Kitcher).

---

148 Joelle Proust (1995) rapporte une reformulation de la définition de Wright inspirée par Manuel García-Carpintero, qui inclut aussi une référence au contexte et rend explicite l'interprétation de la définition en termes de types : « *La fonction d'une occurrence x de type* X *est* Z *ssi : (1') Toutes choses égales par ailleurs, le processus de type* Z *est dans un contexte causal* C *une conséquence typique (résulte) de l'existence d'une structure de type* X *; et (2')* x *existe parce que, toutes choses étant égales par ailleurs, le processus de type* Z *est dans un contexte causal* C *une conséquence typique (résulte) de l'existence d'une structure de type* X. »

L'un des avantages de la définition téléologique par rapport à celle de Wright est qu'elle évite les contre-exemples où l'une des conséquences de la présence de *X* provoque son auto-sélection sans pour autant être sa fonction, comme dans l'exemple de la fuite de gaz toxique. Certains auteurs (Neander, 1991b) évitent ce problème en limitant l'application de la seconde condition de Wright au cadre de la théorie darwinienne : *X* existe parce qu'il a été naturellement « sélectionné pour » $f_{(X)}$. D'autres (Bedau, 1992b) introduisent une dimension valorative dans la définition : une conséquence de la présence de *X* ne peut être sa fonction que si cette conséquence est bénéfique pour *S* ou pour quelqu'un d'autre (nous y reviendrons plus en détail au CHAP. XIII). La définition téléologique inclut cette dernière intuition à travers la fin $F_{(S)}$ sans faire explicitement référence à une valeur. Dans l'exemple de la fuite de gaz, le fait de tuer les personnes qui essaient de la réparer n'est pas l'une de ses fonctions, car cette conséquence de la fuite ne contribue à aucune fin du système, à moins qu'il ne s'agisse d'un sabotage par exemple.

À partir des fonctions étiologiques de Wright on peut retrouver les fonctions darwiniennes qui en dérivent. Si l'on interprète le terme $F_{(S)}$ de la définition téléologique comme étant la *fitness* de l'organisme dans son ensemble ou la fin d'un sous-système contribuant à cette *fitness*, on obtient soit une définition propensionniste soit une définition historique selon le contexte *C* dans lequel on se place. Si le contexte *C* correspond à l'environnement dans lequel un trait *X* est apparu et s'est diffusé dans une population, alors $f_{(X)}$ est la fonction de *X* ssi (i) $f_{(X)}$ est une conséquence de *X* ayant contribué à la *fitness* de l'organisme dans cet environnement, et (ii) $f_{(X)}$ fait partie du système fonctionnel *S* qui définit cet organisme. Étant donné que *X* est un type de trait dans un type d'organismes *S* pendant une période historique *C*, la conséquence $f_{(X)}$ en question est forcément typique elle aussi et pas accidentelle. On retrouve ainsi une attribution fonctionnelle cohérente avec la théorie étiologique faible de Buller (1998). Et si l'on y ajoute une condition supplémentaire consistant à dire que la contribution de $f_{(X)}$ à la *fitness* de l'organisme n'est pas suffisante et qu'il doit y avoir eu également « sélection pour » $f_{(X)}$, on obtient alors une attribution fonctionnelle cohérente avec la conception étiologique dite « forte » au sens de Buller :

> La/une fonction d'un *token* de type *X* dans un système de type *S* et dans le contexte *C* est $f_{(X)}$ si :
> (i) $f_{(X)}$ est une conséquence de *X* qui a été sélectionnée dans *C* et
> (ii) $f_{(X)}$ appartient à *S* dans *C*.

L'une des différences entre cette formulation et celle de Neander est que l'objet de la sélection n'est pas le trait lui-même (*X*) mais l'une de ses conséquences ($f_{(X)}$). Le fait que *X* soit le cœur tel que nous le connaissons et pas un objet différent remplissant les mêmes fonctions n'a ici aucune

importance. Bien sûr, le fait que le cœur soit tel que nous le connaissons et pas autrement revêt une grande importance dans l'économie de l'organisme dans son ensemble, car s'il était possible de le remplacer par un autre système de pompe cela aurait inévitablement des conséquences collatérales ; mais quand on se limite à considérer *une* fonction de *X*, alors la nature de l'objet remplissant cette fonction est contingente et un cœur artificiel fait aussi bien l'affaire qu'un organe naturel.

Une autre différence entre cette formulation et celles de Neander et de Millikan est que : si l'explication fonctionnelle à laquelle nous prétendons n'est pas une explication causale de la présence du trait mais plutôt une justification rationnelle de cette présence, alors peu importe que le *token* auquel on attribue la fonction appartienne à la première génération des porteurs du trait ou au contraire à un lointain descendant. En effet, les réticences à l'égard des premières générations tiennent au fait que les origines causales d'un nouveau trait sont, pour simplifier, aléatoires, la sélection naturelle ne pouvant opérer que sur un trait préexistant (c'est-à-dire existant avant d'avoir été sélectionné). Pour nous, la présence d'un *token* est justifiée, indépendamment de son origine causale, en vertu de son appartenance à un type de trait dont on sait rétrospectivement que l'une de ses conséquences a été sélectionnée : si l'on sait que la capacité qu'a la chlorophylle (*X*) de réaliser la photosynthèse ($f_{(X)}$) a été historiquement sélectionnée (chez les organismes *S* et dans le contexte *C*), alors on pourra attribuer à toutes les molécules de chlorophylle (dans les organismes *S* et le contexte *C*) la fonction de réaliser la photosynthèse ($f_{(X)}$) sans se préoccuper de savoir si une molécule donnée appartient ou pas aux premières générations.

3. Application à plusieurs exemples problématiques

La définition relationnelle de Walsh parle de contribution à la « *fitness* individuelle moyenne » et précise que cette contribution doit être « positive (et significative) ». Cela lui permet d'assurer la distinction chère à Larry Wright entre conséquences fonctionnelles et conséquences accidentelles.

Dans la définition téléologique que nous proposons, cette distinction est assurée par la condition (iii) stipulant que $f_{(X)}$ doit appartenir au système *S* (dans le contexte *C*). En effet, lorsque l'on réalise une analyse fonctionnelle d'un système, les contributions accidentelles au fonctionnement de ce système sont exclues. Dans l'exemple de la boucle de ceinturon de Wright, l'analyse fonctionnelle du « système de protection », s'il existe, révèle que la boucle n'en fait pas partie puisqu'elle n'est pas faite ni utilisée pour se protéger contre les balles (voir note 136, p. 241).

Le système S, la fin $F_{(S)}$ et le contexte C sont interdépendants. Dans le cas du mécanisme d'Anticythère comme dans celui de la boucle de ceinturon, la détermination du système dont on fait l'analyse dépend de l'identification ou de la formulation d'une hypothèse concernant la fin $F_{(S)}$, laquelle doit s'ajuster à son tour aux éléments matériels dont on dispose et au contexte. Dans le cas de la prétendue batterie babylonienne (voir p. 136), le contexte historique ne permet d'imaginer aucun usage possible du très faible courant électrique qu'il aurait été possible de générer avec le vase en question, de sorte qu'il n'est pas raisonnable de lui attribuer comme fin $F_{(S)}$ une production électrique dont les babyloniens n'auraient vraisemblablement tiré aucun bénéfice et dont rien ne nous donne à penser qu'ils connaissaient l'existence. Dans le cas des montres dont la trotteuse sert à enlever la poussière (voir p. 241), c'est au contraire la conception de l'objet, c'est-à-dire l'organisation du système, qui donne une indication de son contexte. Étant donné la finalité générale d'une montre, qui est d'indiquer l'heure, un horloger d'un pays tiers pourrait découvrir la fonction seconde de la trotteuse des montres de P en s'interrogeant sur les raisons ayant pu motiver la singularité de leur conception. Nous disions plus haut que si elles ont été conçues pour enlever la poussière, elles ont sans doute été optimisées pour cela ; donc, dans la mesure où l'analyse du système (c'est-à-dire des éléments matériels en présence) permet de savoir ce pour quoi elles sont optimisées, elle permet aussi de découvrir les raisons et le contexte de leur conception.

Le même raisonnement est applicable à la biologie depuis une perspective créationniste, les êtres vivants y étant conçus comme des artefacts. La définition que nous proposons peut ainsi rendre compte du type d'attribution fonctionnelle que l'on imagine chez Harvey où la découverte de la fonction du cœur est directement liée à celle de la circulation sanguine avec l'unification des systèmes artériel et veineux : le cœur X appartient à un unique circuit fermé S (relation partie-tout) dont la fin $F_{(S)}$ est la circulation du sang et à laquelle il contribue en faisant office de pompe (relation moyens-fin)[149]. Sa découverte explique à la fois le fonctionnement du système circulatoire à travers le rôle causal des éléments qui le composent et la présence du cœur dans l'organisme en vertu de ses conséquences pour la circulation et, plus généralement, en vertu de sa contribution à la fin $F_{(S)}$ assignée par son créateur au système circulatoire :

149 Bien que Harvey ait effectivement découvert le système circulatoire et le rôle moteur du cœur dans ce système, la fonction qu'il lui attribuait ne se limitait pas à favoriser le mouvement du sang mais conservait également celle que lui attribuaient ses prédécesseurs, celle de chauffer, pour liquéfier et vivifier le sang et ainsi le protéger de la corruption et de la coagulation, et aussi pour chauffer le reste du corps (Halleux, 1998).

> « Harvey's claim that the function of the heart is to pump the blood can be understood as proposing that the wise and beneficent designer foresaw the need for a circulation of blood, and assigned to the heart the job of pumping. » (Kitcher, 1993, p. 259)

4. Libéralité apparente de la définition

La libéralité apparente de la réponse peut donner à penser que n'importe quel trait est potentiellement fonctionnel dans un contexte propice. Cette critique rejoint celle formulée à l'encontre de la définition dispositionnelle de Cummins. Mais les choses ne sont pas aussi simples, car pour attribuer une fonction au bruit des battements cardiaques, il faut aussi identifier un système auquel il est censé appartenir et une fin de ce système. Le prétendu système sonore de l'organisme qu'évoque Cummins dans sa réponse à la critique n'a aucune finalité apparente ou, du moins, aucune finalité rationnellement justifiable. Il est vrai que le bruit des battements cardiaques est utile pour prévenir et traiter certaines maladies et contribue par conséquent à la survie des personnes affectées. Cependant, cette contribution requiert la participation d'éléments externes à l'organisme, à savoir, du personnel et des techniques médicales appropriées. Le système *S* à travers lequel le bruit des battements cardiaques contribue aux bien-être de l'organisme est donc en grande partie extra-organique. Sa fin $F_{(S)}$ est la prévention et le traitement des maladies. Son contexte *C* est celui où existent les connaissances, compétences et moyens nécessaires pour le diagnostic et le traitement des maladies en question. Par conséquent, s'il est possible d'attribuer une fonction au bruit des battements cardiaques, c'est seulement en tant qu'élément d'un système technique, une fonction comparable à celle d'une partie d'un artefact plutôt qu'à ce qu'on entend habituellement par une fonction biologique.

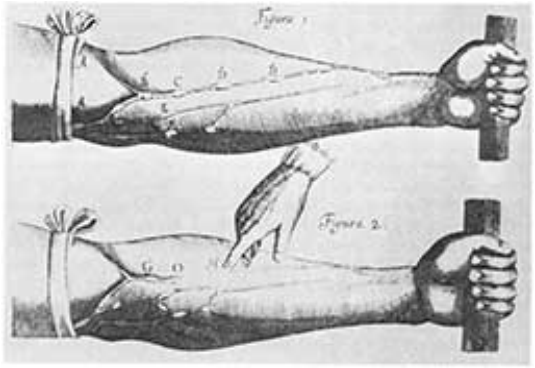

Figure 22: Circulation sanguine. Mise en évidence par Harvey du sens du mouvement du sang dans les veines, dans l'Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus. (Source : Wikipedia)

On peut déterminer le système auquel appartient un élément fonctionnel en examinant les conséquences de sa délétion. Si, comme dit Wittgenstein (1984, paragr. 271), « *une roue qui tourne sans que rien ne*

bouge avec elle ne fait pas partie du mécanisme »[150], alors le fait d'arrêter la roue peut aider à connaître le mécanisme auquel elle appartient. Par exemple, en liant les membres à l'aide d'un lacet pour y interrompre le flux sanguin, Harvey observe que le sang va du cœur vers la périphérie dans les artères, et de la périphérie vers le cœur dans les veines (Fig. 22). De cette manière, et en réalisant un calcul quantitatif du volume de sang chassé par le cœur dans l'aorte à chaque pulsation, Harvey en vient à l'idée d'un mouvement en circuit fermé, ce qui implique que le système artériel et le système veineux n'en font en réalité qu'un seul dont le cœur constitue à la fois le moteur et le nœud[151]. De plus, on sait aujourd'hui qu'une entrave à la circulation sanguine (par garrottage, clampage, thrombose, compression vasculaire ou hémorragie) provoque à terme la nécrose par anoxie des tissus affectés, ce qui met en rapport la circulation sanguine avec le métabolisme cellulaire, d'une part, et la respiration, de l'autre. Par ailleurs, on sait aussi que la réduction du débit sanguin liée à une insuffisance cardiaque congestive déclenche une série de mécanismes de compensation neuro-hormonaux visant à accroître ce débit en agissant sur les éléments qui le composent : le cœur, par augmentation de la fréquence de contraction et par remodelage ventriculaire ; les vaisseaux sanguins, par vasoconstriction ; le liquide transporté, par rétention d'eau et de sel. À titre de comparaison, la réduction ou l'interruption du bruit des battements cardiaques, toutes choses étant égales par ailleurs, n'aurait aucune conséquence pour le reste de l'organisme. Nous avons donc de bonnes raisons de penser que ce bruit, comme la roue qui tourne à vide, ne fait pas partie du mécanisme.

Le bruit des battements cardiaques est une conséquence de la présence du cœur, mais elle ne contribue pas au fonctionnement de l'organisme ni, en dernière instance, à sa survie et sa reproduction, et quand bien même elle y contribuerait, puisqu'elle peut effectivement servir à diagnostiquer une maladie, cette contribution serait accidentelle, car elle

150 La réflexion de Wittgenstein porte sur la signification et la possibilité d'un langage privé, ce qui n'a *a priori* rien à voir avec les fonctions. Pour autant, le rapprochement n'est pas fortuit car les fonctions, d'après nous, relèvent moins du fait biologique (ce que Searle appelle un « fait brut ») que de l'interprétation que l'on en fait et de la signification qu'on lui donne.

151 Pierre Flourens illustre avec une expérience on ne peut plus graphique son commentaire de la découverte par Harvey de la circulation du sang : « *Dans mes leçons au* Jardin des Plantes*, pour simuler, sous les yeux de mes élèves, le passage du sang des artères dans les veines, je fais l'expérience suivante : Je fais ouvrir, sur un chien mort, l'artère et la veine crurales. On insère ensuite une canule dans le bout ouvert de l'artère, et on pousse de l'eau au moyen d'une seringue. Au bout de très-peu d'instants, l'eau, injectée par l'artère, revient par la veine. C'est l'image complète de la circulation.* » (Flourens, 1857, p. 49 note 1).

ne fait pas partie du système organique. On ne peut donc pas lui attribuer de fonction biologique.

Peut-on néanmoins lui attribuer une fonction technique dans le contexte de la médecine moderne ? Si les artefacts s'inscrivent dans des systèmes sociaux (Krohs, 2008), la portée des fonctions techniques s'en trouve élargie. Dès lors, on pourrait penser que le bruit des battements cardiaques fait partie d'un système technique et social comprenant des artefacts et des pratiques médicales dont la finalité est la prévention, le diagnostic et le traitement des maladies. Depuis cette perspective, s'il n'a pas de fonction biologique, le bruit des battements pourrait malgré tout avoir une fonction technique.

La réponse est négative. En effet, on peut dire que la médecine utilise le bruit du cœur pour établir un diagnostic, mais on ne peut pas dire que le cœur fasse du bruit pour rendre possible ce diagnostic. De même, les lunettes utilisent la protubérance nasale pour se maintenir en place, mais les nez ne sont pas saillants pour supporter les lunettes. Les plantes utilisent l'eau de pluie pour croître, mais la pluie ne tombe pas pour arroser les plantes. Le cœur, le nez, la pluie sont indépendants des systèmes qui les exploitent. Ils n'appartiennent pas au système *S*, mais à ses conditions de fonctionnement *C*. Conformément à la troisième condition de la définition, on ne peut donc pas leur attribuer la fonction en question.

Conclusions de la troisième partie

Une fonction peut être caractérisée en première approximation comme la contribution d'un élément à une fin du système auquel elle appartient. En effet, une relation fonctionnelle est à la fois une relation entre une partie et un tout, et entre un moyen et une fin.

Par ailleurs, une même fonction peut être réalisée par des structures et des processus matériels différents. Il convient donc de distinguer entre la fonction et la structure matérielle qui lui sert de support. Or, un système fonctionnel n'est pas un ensemble de structures mais un ensemble de fonctions. L'analyse fonctionnelle d'un tel système se situe donc à un niveau d'abstraction plus élevé que l'analyse de sa composition et organisation matérielle.

Une attribution fonctionnelle est à la fois une explication de la présence d'un trait dans un système donné, à la manière de Wright, et une explication causale des capacités de ce système, à la manière de Cummins. Mais nous défendons une explication téléologique c'est-à-dire non causale, de la présence du trait où la finalité du système justifie les moyens de sa réalisation. Et tandis que l'analyse fonctionnelle de Cummins s'en tient à l'organisation matérielle d'un système, ce qui ne lui permet pas d'attribuer une fonction aux fragments rouillés du mécanisme d'Anticythère, nous la situons quand à nous à un niveau d'abstraction plus élevé.

En appliquant l'analyse fonctionnelle à des types plutôt qu'à des *tokens*, on peut par exemple expliquer la capacité du système circulatoire des vertébrés à faire circuler le sang grâce à une pompe qu'on appelle le cœur. Tous les vertébrés ont donc un cœur dont la fonction dans le système circulatoire est de pomper le sang, bien qu'un cœur particulier, c'est-à-dire une masse de tissus concrète, soit incapable de le faire. La même analyse nous permet d'expliquer pourquoi cet organe est présent chez les vertébrés : parce que leur organisme a besoin de faire circuler le sang, parce que la circulation est assurée par le système circulatoire et parce que le système circulatoire comprend une pompe qui est le cœur. Autrement dit : tels qu'ils sont constitués, tous les vertébrés ont besoin d'un cœur pour faire circuler le sang ; et s'ils n'en avaient pas besoin, ils seraient constitués autrement et feraient l'objet d'une analyse différente.

C'est donc l'organisation du système qui justifie la présence des traits qui le composent dans la mesure où ils contribuent à une fin de ce système.

L'ingénierie inverse permet de distinguer entre une fonction et une conséquence bénéfique accidentelle sans recourir ni à l'histoire ni à la sélection. C'est l'organisation du système — son *design* — qui permet de les distinguer. On peut de cette manière éviter les contre-exemples qui invalident la définition de Wright et proposer une conception des fonctions commune aux êtres vivants et aux artefacts. Cependant, l'ingénierie inverse s'appuie sur un critère d'optimalité dont l'usage en biologie, dans le cadre de l'adaptationnisme, a fait l'objet de nombreuses critiques. Tout en reconnaissant la pertinence de certaines de ces critiques, nous estimons qu'elles ne sont pas rédhibitoires et que la présomption d'optimalité reste un bon critère d'attribution fonctionnelle par défaut, aussi bien en biologie que pour les artefacts.

La définition générale du concept de fonction que nous proposons est la suivante :

> La/une fonction d'un *token* de type X dans un système de type S et dans le contexte C est $f_{(X)}$ si :
> 1. $f_{(X)}$ est une conséquence de X ;
> 2. $f_{(X)}$ contribue à une fin du système $F_{(S)}$;
> 3. $f_{(X)}$ appartient à S.

Cette définition requiert une justification raisonnée de l'attribution d'une fin à un système naturel. C'est ce que nous allons faire dans les deux dernières parties de ce travail.

Quatrième partie :
JUSTIFICATION DE LA TÉLÉOLOGIE

[D]ans l'être organisé et vivant, l'organisation et la vie jouent simultanément le rôle d'effet et de cause, par une réciprocité de relations qui n'a d'analogues, ni dans l'ordre des phénomènes purement physiques, ni dans la série des actes soumis à l'influence d'une détermination volontaire et réfléchie

Antoine Augustin Cournot, *Essai…*, 1851, IX, § 129

Cette détermination intentionnelle paraît surtout évidente dans les êtres vivants qui forment un tout fini ; elle le paraît moins pour le physicien et le chimiste qui ne voient que des fragments des phénomènes généraux du grand tout. Aussi sont-ce ceux-ci qui ont combattu la téléologie comme fournissant des idées fausses et aujourd'hui les savants n'osent pas avouer qu'ils sont téléologistes parce que ce sont des choses qui ne se démontrent pas

Claude Bernard, *Cahier de notes* (1860, p. 59)

C'est la signification qui est le fil directeur sur lequel la biologie doit se guider, et non la misérable règle de causalité qui ne peut voir plus loin qu'un pas en avant ou un pas en arrière, et reste aveugle aux grandes relations structurelles.

Jakob von Uexküll,
Mondes animaux et monde humain (1956, p. 117)

Introduction de la quatrième partie

La définition du concept de fonction que nous avons proposée au chapitre précédent dépend directement de l'attribution de fins aux systèmes contenants. Dès lors, il convient de répondre à deux questions que nous avons posées de façon récurrente tout au long de ce travail, à savoir : *Pourquoi attribue-t-on des fonctions et des fins à certaines choses mais pas à d'autres ?* et *Comment justifier ces attributions dans un cadre scientifique ?*

Nous commencerons par la première au CHAP. VIII. Nous essaierons de comprendre, d'un côté, pourquoi nous avons une tendance apparemment spontanée à penser le monde en termes téléologiques et, de l'autre, pourquoi certaines choses s'y prêtent et d'autres pas. La question a donc une dimension subjective, puisqu'elle porte sur une disposition du sujet, et une dimension objective portant sur les caractéristiques de l'objet. Nous l'aborderons depuis la psychologie cognitive pour essayer d'en dégager à la fois la portée, les origines et les conditions de possibilité.

La pensée téléologique ou téléofonctionnelle est-elle effectivement sélective, c'est-à-dire limitée à certains domaines comme la biologie, ou au contraire s'applique-t-elle de manière générale à tous les objets possibles ? Est-elle innée ou acquise, propre à l'humain ou partagée par d'autres animaux, primitive ou dérivée d'autres mécanismes mentaux ?

En particulier, nous chercherons à savoir si elle est liée d'une manière ou d'une autre aux pensées causale et intentionnelle, car plusieurs des critiques qui lui ont été adressées s'appuient sur ces liens supposés. D'une part, on accuse la téléologie d'être une forme invalide de la causalité, soit parce qu'elle en serait une forme immature, soit parce qu'elle la détournerait de son usage correct, par exemple sous la forme de causes finales ou rétrogrades. D'autre part, on l'accuse d'anthropomorphisme, d'artificialisme, de finalisme ou encore d'animisme, c'est-à-dire de supposer ou de projeter une intentionnalité dans les phénomènes naturels. Si la téléologie était psychologiquement indépendante de la causalité et de l'intentionnalité, ces critiques y perdraient beaucoup de leur force.

Nous examinerons ensuite la seconde des deux questions. Nous nous demanderons aux CHAP. IX et X s'il est scientifiquement acceptable d'attribuer des fins à des objets naturels non intentionnels, et, au CHAP. XI, quelle signification et quelle valeur scientifique ces attributions peuvent éventuellement avoir en biologie.

Pour certains critiques, la téléologie est fausse, tout simplement, car elle s'appuie sur des croyances métaphysiques qui n'ont pas leur place dans les sciences ou sur des théories pseudo-scientifiques obsolètes. Pour d'autres, c'est un mode d'expression et de pensée peu rigoureux, une façon de parler qu'il convient d'éviter car elle prête à confusions. La plupart des auteurs que nous avons cités dans ce travail estiment quant à eux que la téléologie est acceptable en biologie et qu'elle a même une valeur explicative dans la mesure où les explications téléofonctionnelles sont en réalité des explications causales déguisées que l'on peut réduire ou traduire, ou expliciter en termes de causes efficientes.

En ce qui nous concerne, nous allons défendre à la fois le caractère non causal de la téléologie et sa légitimité scientifique. Tout d'abord, nous réfuterons certaines des critiques les plus fréquentes en montrant, à partir d'exemples simples, comment les explications téléologiques fonctionnent effectivement. Ensuite, nous montrerons que la téléologie biologique est indépendante de la nature physique de ses objets et nous réfuterons d'autres objections courantes. Finalement, nous nous pencherons sur les contributions positives que la pensée téléofonctionnelle apporte à la biologie et qui seraient difficiles ou impossibles à obtenir autrement. Nous verrons en effet qu'elle permet de réaliser des prédictions et des catégorisations extrêmement importantes, et que ses explications ont une vraie valeur explicative.

Chapitre VIII

Pourquoi la téléologie est-elle sélective ?

Selon Jean Piaget (1947), la représentation du monde chez l'enfant se caractérise par trois grandes lignes : le réalisme, l'animisme et l'artificialisme. Le réalisme de l'enfant est une confusion entre l'interne et l'externe, le psychique et le physique ; c'est la tendance à situer dans les choses ce qui est un produit de l'activité du sujet pensant. L'animisme est la tendance à considérer les choses comme vivantes et conscientes. Il traverse plusieurs étapes au cours desquelles l'enfant va d'abord attribuer vie et conscience à tout ce qui a une activité (évaluée en fonction de son utilité pour les personnes), puis à tout ce qui est en mouvement (par opposition à ce qui est inerte), puis aux corps doués de mouvement propre (par opposition au mouvement acquis), et enfin aux animaux seuls. L'artificialisme est la tendance à expliquer l'origine des objets et des phénomènes naturels de la même manière que les produits de l'intelligence humaine, c'est-à-dire comme s'ils étaient l'œuvre de l'être humain ou d'un être agissant comme lui. Par exemple, le soleil est conçu comme « un caillou tout rond qu'un monsieur a allumé puis lancé dans le ciel ». Il convient de noter que cet artificialisme est compatible avec l'animisme, puisqu'une même chose peut être à la fois fabriquée et vivante, et les enfants considèrent en général la naissance des bébés comme une fabrication. Par ailleurs, il convient aussi de noter que l'une des racines de l'artificialisme est le finalisme : la nature est faite *par* nous dans la mesure où elle est faite *pour* nous. Ainsi, le soleil « c'est pour nous chauffer », la pluie « c'est pour arroser », les montagnes « c'est pour aller se promener », etc.

L'artificialisme serait une forme primitive de causalité qui évolue à mesure que l'enfant grandit et qui laisse la place à une explication de la nature par elle-même, sans faire appel à l'activité humaine[152]. Il serait suivi d'une rationnalisation progressive de la représentation du monde conduisant à une compréhension pleinement causale, au sens de la causalité physique (efficiente). Par conséquent, l'attribution de fins aux choses extérieures correspondrait à un mode de pensée par défaut dont la portée se réduit au fur et à mesure que l'enfant apprend à remplacer les explications téléologiques par des explications causales. Mais s'il en est ainsi, alors pourquoi l'animisme et le créationnisme sont-ils présents dans la pensée religieuse, et pourquoi les concepts et les explications téléologiques sont-ils employés par la biologie actuelle ? Doit-on en conclure que ce sont des rémanences du raisonnement enfantin chez l'adulte ?

Certaines des conclusions de Piaget, comme l'animisme des enfants préscolaires, ont été confirmées par de nombreuses études postérieures (voir Carey, 1985, Chapitre 1). D'autres ont été contestées. En particulier, l'idée selon laquelle les enfants de 3 à 7 ans ne seraient pas capables de raisonner en termes de causes physiques a été battue en brèche par des études qui montrent que les nourrissons peuvent percevoir des relations de cause à effet dès l'âge de 6 mois (Leslie, 1982; Leslie & Keeble, 1987) et qu'ils développent leur compréhension du monde physique au cours des mois précédents (Baillargeon & DeVos, 1991; Baillargeon, Spelke, & Wasserman, 1985; Luo & Baillargeon, 2005; Rosander & von Hofsten, 2004; Ruffman, Slade, & Redman, 2005). On ne peut donc pas considérer la téléologie comme une forme de pensée précausale qui s'appliquerait aux objets naturels à défaut d'une autre explication plus mature.

1. Y a-t-il une biologie naïve autonome ?

La question du rapport entre la pensée téléologique et la représentation du monde chez l'enfant demeure pertinente aujourd'hui. Les chercheurs en psychologie du développement cognitif s'intéressent particulièrement au rôle que jouent les attributions de fins dans la construction d'une théorie naïve de la biologie. L'un des aspects les plus polémiques de cette question concerne l'existence chez l'enfant d'une conception du domaine biologique qui serait indépendante de leur représentation des animaux comme des êtres intentionnels. En d'autres termes, la biologie naïve de l'enfant est-elle autonome de sa psychologie naïve ? Paul Bloom (2004), par exemple, défend l'idée que les petits enfants sont naturellement dualistes, c'est-à-dire qu'ils voient et comprennent de manière différente les corps et les esprits, le monde

152 Antonio Battro (1966) distingue quatre périodes, la dernière allant jusqu'à l'âge de 9-10 ans.

physique et le monde social, de sorte qu'à l'âge de 5 mois ils ne considèrent pas les humains comme des objets matériels (Kuhlmeier, Bloom, & Wynn, 2004) et qu'ils attribuent des états mentaux (désirs, croyances, intentions) à n'importe quel objet dont le comportement montre des signes d'agentivité (Premack & Premack, 1997).

Ces questions sont pertinentes pour nous dans la mesure où deux des principales critiques formulées contre la téléologie biologique consistent à dire qu'elle est incompatible avec la physique, puisqu'elle impliquerait une causalité rétrograde, et qu'elle est anthropomorphe, puisqu'elle impliquerait l'attribution d'états mentaux à des objets qui n'en ont pas (voir Mayr, 1988, Chapitre 3). Dès lors, s'il existait une pensée biologique autonome de la pensée physique et si la téléologie ne dépendait pas de l'intentionnalité, alors ces critiques n'auraient plus beaucoup de sens.

À ce jour, malheureusement, il n'y a pas de consensus sur ces questions. En ce qui concerne la première, les avis sont partagés entre ceux qui admettent l'existence de modules cognitifs spécifiques au domaine biologique (Atran, 1995), ceux qui estiment que les enfants développent très tôt une biologie naïve sur la base d'une pensée téléologique autonome, sélective et innée (Inagaki & Hatano, 2002, 2006; Keil, 1992), et ceux qui estiment que la pensée téléologique n'est pas autonome ni limitée à un domaine et ne devient sélective qu'avec l'âge et l'éducation (Carey, 1985; Kelemen, 1999b).

1.1. Ceux qui sont contre

À la suite de Piaget, Susan Carey (1985) a étudié l'évolution de plusieurs concepts biologiques chez l'enfant et en a conclut qu'un changement se produit vers l'âge de 10 ans. D'après elle, les enfants d'âge préscolaire n'ont pas de connaissance spécifique au domaine biologique et raisonnent à propos des animaux à partir d'une « psychologie naïve » centrée sur la capacité à agir — où l'action est entendue en termes de causes intentionnelles —, et sur une « biologie vitaliste » centrée sur le mouvement propre ou l'activité auto-générée. À partir de l'âge de 10 ans, les enfants manifestent ce qu'elle appelle une « biologie mécaniciste », c'est-à-dire une conception de l'organisme entendu comme une machine composée de systèmes interconnectés qui contribuent à son maintien et à son fonctionnement (Carey, 1988).

Deborah Kelemen (1999b, 1999a) affirme quant à elle que la téléologie n'est pas initialement limitée mais s'applique de façon générale à tous les objets et ne devient sélective que plus tard au cours du développement. Partant du constat qu'il existe un lien étroit entre le raisonnement fonctionnel et intentionnel qui perdure chez l'adulte, elle a réalisé une série d'études comparatives qui tendent à montrer que :

1. les enfants attribuent des fonctions à tous les objets, sans limitation de domaine ;
2. ils interprètent la fonctionnalité en termes d'intention originale (« créé pour ») ; et
3. bien qu'ils attribuent une fonction à davantage d'objets que les adultes, ils partagent la même notion de fonction, entendue comme « *original design* ».

L'auteure en conclut qu'enfants et adultes diffèrent quand à leur connaissance de l'histoire causale des objets : les premiers conçoivent les phénomènes naturels de la même manière que les artefacts, c'est-à-dire comme étant créés par un agent non-humain. En l'absence de connaissances quant à l'histoire causale d'un objet (ses origines), ils construisent celle-ci en termes intentionnels, qui sont plus proches de leur propre expérience. Une autre étude montre que pour expliquer pourquoi un animal a un long cou ou pourquoi des rochers sont pointus, les enfants de moins de 10 ans préfèrent les explications téléologiques et intentionnelles face aux explications physiques (1999c). Le raisonnement téléofonctionnel serait donc un mode de raisonnement par défaut en l'absence de connaissances scientifiques qui s'atténue à mesure que l'on acquière une meilleure compréhension causale (physique) du monde naturel mais qui demeure prégnant chez l'adulte (voir Casler & Kelemen, 2008; Kelemen & Rosset, 2009; Kelemen, Rottman, & Seston, 2012).

Les conclusions de Kelemen rejoignent-elles celles de Piaget ? Le savant suisse croyait que la pensée téléologique était une pensée immature qui précédait la pensée physique. Or, nous savons aujourd'hui que ce n'est pas le cas, car les deux sont disponibles très tôt au cours du développement cognitif. Il s'agirait plutôt de modes de raisonnement alternatifs entre lesquels nous pouvons choisir pour expliquer, catégoriser et prédire les phénomènes du monde extérieur. La question est de savoir pourquoi nous employons l'un plutôt que l'autre pour certains objets et pas pour d'autres. D'après Déborah Kelemen, c'est l'éducation moderne occidentale qui nous pousse à choisir de préférence les explications physiques pour les phénomènes naturels non-biologiques et à limiter les attributions téléofonctionnelles aux artefacts et aux traits biologiques. Les pensées de l'enfant et de l'adulte ne seraient donc pas incommensurables. En ce qui concerne l'artificialisme, une étude de Kelemen et DiYanni (2005) montre que les enfants d'âge scolaire ont tendance à expliquer tous les objets à la manière des artefacts, mais elle montre aussi que l'adoption du dessein intelligent n'est pas aussi large, car ils l'appliquent de préférence à l'origine des animaux et des artefacts.

1.2. Ceux qui sont pour

Pour d'autres auteurs, les enfants ont une pensée proprement biologique — non-intentionnelle — à un âge précoce. Selon Frank Keil (1992), la biologie naïve dépendrait directement de la tendance téléologique innée à voir certaines entités comme *conçues pour* quelque chose. Par exemple, on montre à des enfants des émeraudes et des plantes, et on leur demande de choisir entre deux explications de leur couleur verte : une explication fonctionnelle (« parce que c'est mieux pour les plantes et cela aide à ce qu'il y en ait plus ») et une explication physique (« parce que les plantes ont des toutes petites parties qui leur donnent cette couleur »). Tandis que les tout-petits enfants ne montrent aucune préférence explicative, ceux âgés de 5 à 7 ans adoptent l'explication téléologique pour les plantes et la mécanique pour les émeraudes. De plus, dès l'âge de 3 ans, les enfants attribueraient des fonctions aussi bien aux artefacts qu'aux traits biologiques, mais de façon différente : tandis que les premiers existent pour servir des agents externes, les seconds existent pour le bien des organismes eux-mêmes (Keil, 1995). Et lorsqu'ils s'interrogent sur des artefacts et des animaux qu'ils ne connaissent pas, les enfants de 3 à 5 ans posent des questions différentes pour les uns et les autres qui trahissent leurs connaissances des propriétés essentielles propres à ces deux domaines d'objets (Greif, Kemler Nelson, Keil, & Gutierrez, 2006). En outre, dès l'âge de 8 mois, ils distingueraient les animaux d'autres objets mobiles et feraient des inférences quant à leur contenu, ce qui semble confirmer l'apparition précoce d'une pensée biologique autonome (Keil, 2013). Ainsi, les petits enfants auraient très tôt à leur disposition plusieurs modèles explicatifs pour penser les phénomènes d'un domaine donné (Gutheil, Vera, & Keil, 1998).

Inagaki et Hatano (2006) estiment eux-aussi que la biologie naïve des enfants de 5 ans ne dépend pas de la psychologie ni de la physique naïves. Ils disposeraient d'une espèce de théorie du domaine biologique qui leur permet de faire des prédictions à propos des organismes et qui repose sur deux composants : (1) la distinction entre le vivant et le non-vivant et entre le corps et l'esprit, et (2) des instruments de pensée spécifiques aux phénomènes biologiques. D'après eux, le premier de ces instruments de pensée est la « téléologie du vivant » (*life teleology*), c'est-à-dire l'idée que les traits biologiques ont une fonction et qu'ils existent pour le bien de l'organisme, comme l'avait montré Keil (1992). L'autre instrument est la causalité entendue en termes de force vitale (*vital power*), c'est-à-dire l'idée que les processus biologiques servent à maintenir le corps en vie grâce à la force vitale qu'ils prennent de l'extérieur. Inagaki et Hatano (2002) avaient en effet montré que face à des phénomènes comme la digestion et la respiration, les enfants de 6 ans optent pour des explications causales vitalistes de préférence aux explications physiologiques et intentionnelles. Par exemple, à la question « pourquoi

inspire-t-on de l'air », la réponse vitaliste est qu'on le fait « parce que nos poumons extraient de la force vitale de l'air ». Une réponse qui n'est ni intentionnelle ni mécanique.

1.3. Pour aller plus loin

Les contradictions manifestes entre les résultats des différentes études ne permettent pas, à l'heure actuelle, de trancher la question de l'existence d'une biologie autonome chez l'enfant.

À défaut d'apporter une réponse à cette question, d'autres études peuvent cependant nous aider à avancer dans la réflexion. Par exemple, des résultats obtenus par imagerie fonctionnelle indiquent que le cerveau pourrait être naturellement « câblé » pour percevoir le mouvement biologique et en extraire des informations permettant l'identification de l'agent qui en est responsable, ce qui s'avère particulièrement utile pour catégoriser ses mouvements comme menaçants ou séduisants et pour prédire ses actions futures (voir Blakemore & Decety, 2001). En effet, en milieu naturel, la survie animale dépend en grande partie de la capacité à distinguer les mouvements biologiques d'autres formes de mouvement, à identifier ceux des prédateurs, des proies et des partenaires sexuels, et à anticiper leurs actions. Cette capacité de discrimination a été mise en évidence chez les nouveaux-nés humains âgés de quelques jours, malgré l'immaturité de leur système visuel (Méary, Kitromilides, Mazens, Graff, & Gentaz, 2007; Simion, Regolin, & Bulf, 2008), et chez les nouveaux-nés d'autres espèces animales, comme les poules (Vallortigara, Regolin, & Marconato, 2005).

Par ailleurs, d'autres études semblent indiquer l'existence d'un biais cognitif téléofonctionnel précoce qui fait que les enfants conçoivent les artefacts en termes fonctionnels dès l'âge de 2 ans (Casler & Kelemen, 2007; Kemler Nelson, Egan, & Holt, 2004) et qu'ils prêtent attention à leurs propriétés fonctionnelles dès l'âge de 1 an (voir Hernik & Csibra, 2009), ce qui n'est pas étonnant si l'on considère que d'autres animaux, comme les corbeaux, les simiens et les anthropoïdes, en sont également capables (Clair & Rutz, 2013; Hauser, 2002; Herrmann et al., 2008; Ruiz & Santos, 2013; Santos, Miller, & Hauser, 2003).

Au-delà du caractère général ou sélectif de ce biais, la question qui nous intéresse est celle du lien qu'il entretient avec la psychologie naïve. L'attribution de fins aux objets naturels implique-t-elle nécessairement l'attribution d'états mentaux extrinsèques (artificialisme) ou intrinsèques (animisme, anthropomorphisme) ? La téléologie est-elle un produit dérivé mais distinct de la psychologie ? S'agit-il au contraire de deux modes de pensée radicalement différents ? Ou est-ce plutôt la psychologie qui dérive de la téléologie ? Pour répondre à ces questions, il faudrait pouvoir remonter aux origines du développement cognitif. Or, la plupart des études précédemment citées s'appuient sur des interactions verbales avec

des adultes et des enfants en âge de parler. D'autres études, méthodologiquement différentes, sont capables d'interroger les nourrissons, mais les questions qu'elles posent portent sur l'interprétation du comportement des objets — ou plutôt sur la compréhension de l'action des agents — et pas sur leur origine ni sur la fonction de leurs parties. Nonobstant, elles permettent d'éclairer les rapports entre la finalité, l'intentionnalité et la causalité physique et elles pourraient nous aider à dévoiler les racines du biais téléofonctionnel.

2. Aux origines de l'interprétation des actions téléologiques

S'il n'y a pas de consensus quant à l'autonomie de la biologie, il n'y en a pas non plus concernant l'origine et la portée des premières interprétations des actions dirigées vers un but. En ce qui concerne la portée, l'un des sujets de discussion est le type d'agent susceptible d'une telle interprétation : s'applique-t-elle seulement aux actions humaines ou inclut-elle aussi d'autres types d'agents (animaux, plantes, artefacts, phénomènes naturels) ? Un autre sujet de discussion est le type d'action : l'interprétation concerne-t-elle seulement les actions familières à l'enfant ou s'applique-t-elle de façon générale à n'importe quelle action qui montre des indices de comportement dirigé vers un but ?

Une étude a évalué les modèles employés par les enfants et les adultes pour prédire le déroulement d'actions téléologiques — à la fois dirigées vers un but et bénéfiques pour l'agent — en fonction du type d'agent impliqué (Opfer & Gelman, 2001). Les résultats montrent que les adultes adoptent un modèle biologique, c'est-à-dire qu'ils réservent généralement leurs prédictions téléologiques pour les actions des animaux et des plantes (Fig. 23). Ils montrent aussi que les enfants préscolaires et les élèves du cours moyen ont davantage tendance que les adultes à prédire des actions téléologiques et qu'ils le font pour n'importe quel domaine d'objets, y compris des artefacts simples (modèle finaliste), mais chez les préscolaires cette tendance est plus forte pour les animaux que pour les autres objets (modèle animal).

	Domaines capables d'une action téléologique			
Modèles	Animaux	Plantes	Machines	Artéfacts simples
Finaliste	x	x	x	x
Complexe	x	x	x	
Biologique	x	x		
Animal	x			

Figure 23: Quatre modèles hypothétiques pour prédire des actions téléologiques. Source : Opfer & Gelman, 2001.

Les auteurs de l'étude interprètent ces résultats en disant que les préscolaires attribuent vraisemblablement des états mentaux aux animaux, tandis que les adultes croient que le fait d'agir téléologiquement est un aspect essentiel du vivant. Mais cette interprétation n'explique pas pourquoi l'attribution d'états mentaux chez les enfants de 5 ans se limite aux animaux, ni pourquoi ils prédisent néanmoins des actions téléologiques pour tous les objets, ni pourquoi ils continuent à le faire à l'âge de 10 ans alors qu'ils n'attribuent plus d'états mentaux aux objets non-humains.

Par ailleurs, on a cru pendant de longues années que l'enfant n'est pas capable de se représenter les états mentaux intentionnels d'autrui avant l'âge de 4 ans, moment à partir duquel il deviendrait capable d'attribuer des croyances fausses et développerait d'une théorie de l'esprit. Or, des études plus récentes indiquent qu'avant la seconde année les enfants comprennent déjà certains aspects de l'action intentionnelle. En effet, ils seraient déjà capables d'attribuer des croyances à l'âge de 18 mois (Senju, Southgate, Snape, Leonard, & Csibra, 2011), et peut-être même à 13 mois (Surian, Caldi, & Sperber, 2007), ils savent si quelqu'un veut ou pas leur donner un jouet à l'âge de 9 mois (Behne, Carpenter, Call, & Tomasello, 2005), et, dès 3-6 mois, ils attribueraient des buts à des agents et seraient capables de prédire leurs actions téléologiques (moyens-fin) dans des contextes non familiers (voir Gergely, 2011). Par ailleurs, d'autres animaux semblent comprendre eux-aussi certains aspects de l'action intentionnelle et téléologique, bien que l'on ne puisse pas encore affirmer ni écarter l'existence chez eux d'une théorie de l'esprit (Call & Tomasello, 2008; Lurz, 2009). Ce décalage a donné lieu à plusieurs tentatives d'explication qui peuvent s'avérer intéressantes pour nous dans la mesure où elles permettent d'éclairer le rapport primordial entre la téléologie et l'intentionnalité.

2.1. Quel rapport entretiennent la téléologie et l'intentionnalité chez le jeune enfant ?

L'approche défendue par Kelemen (1999a), appelée « *Promiscuous Teleology* », affirme que le raisonnement téléofonctionnel dérive de notre capacité, en tant qu'animaux sociaux, à attribuer des intentions à d'autres agents ; une capacité qui se développe très tôt chez l'humain et qui présente une grande valeur adaptative parmi les primates. En effet, dès l'âge de 1 an, les enfants semblent être capables de comprendre l'utilisation intentionnelle d'un objet par un agent et comment des artefacts (familiers ou pas) peuvent être employés pour atteindre un but. Si on ajoute à cela, dit Kelemen, que les jeunes enfants vivent entourés d'artefacts dont la présence s'explique par la façon dont les agents les utilisent pour atteindre leurs propres fins, on peut supposer que ces premières expériences jouent un rôle crucial dans l'élaboration de mécanismes explicatifs plus généraux. Elles pourraient dès lors contribuer à la tendance persistante, y compris chez l'adulte, à expliquer toutes sortes de choses en termes téléofonctionnels. Autrement dit, à défaut d'une meilleure explication, ils s'appuieraient sur leur connaissance privilégiée des intentions et des artefacts pour expliquer les objets et les phénomènes naturels et en déduire qu'ils existent parce qu'un agent les a créés intentionnellement.

D'autres théories s'appuient sur la découverte — d'abord chez les singes, puis chez l'humain et aussi chez certains oiseaux — de neurones dits « miroir » qui s'activent lorsque l'on voit ou entend quelqu'un réaliser une action (Rizzolatti, 2005; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). D'après certains chercheurs, ces neurones permettraient d'expliquer en termes de processus non-inférentiels, non-représentationnels et non-mentalistes la capacité qu'ont les nourrissons de comprendre certains aspects de l'action intentionnelle (voir Gergely, 2011; et aussi A. I. Goldman, 2012). Il s'agirait en effet d'un système de simulation motrice des actions perçues qui permettrait de s'identifier à l'autre et de saisir en première personne la finalité de ses actions. Ce mécanisme neuronal, automatique et inné, pourrait être le fondement de l'apprentissage par imitation de nouvelles actions de type moyens-fin. Et il pourrait aussi contribuer au développement d'une théorie de l'esprit. Ainsi, contrairement à la théorie de Kelemen, où la téléologie dérive de l'intentionnalité, les neurones miroir constitueraient un mécanisme de simulation qui précède et qui rend possible à la fois la compréhension des actions intentionnelles et des relations moyens-fin. Cependant, toutes ces interprétations des neurones miroir sont encore spéculatives, de sorte que nous ne pouvons pas en tirer de conclusions fermes pour notre réflexion.

2.2. La théorie de l'attitude téléologique

Une troisième approche, appelée « *Teleological Stance* », défend l'existence de deux systèmes innés indépendants : l'un spécialisé dans la représentation et l'interprétation des actions instrumentales (téléologiques), l'autre dans la représentation et l'interprétation des actions communicatives (intentionnelles) (Gergely, 2011). En ce qui concerne le premier, Gergely Csibra, György Gergely et d'autres auteurs ont d'abord montré, en utilisant le protocole de violation des attentes, que les enfants de 9 et 12 mois comprennent les actions dirigées vers un but (Csibra & Gergely, 1998; Csibra, Gergely, Bíró, Koos, & Brockbank, 1999; Gergely & Csibra, 1997; Gergely, Nádasdy, Csibra, & Biro, 1995; A. L. Woodward, 1998). Dans l'une des expériences, on habitue les enfants à voir sur un écran d'ordinateur une animation montrant un petit cercle s'approcher d'un grand cercle en « sautant » par-dessus un rectangle situé entre les deux ; ensuite, on retire le rectangle et on observe leur réaction face à

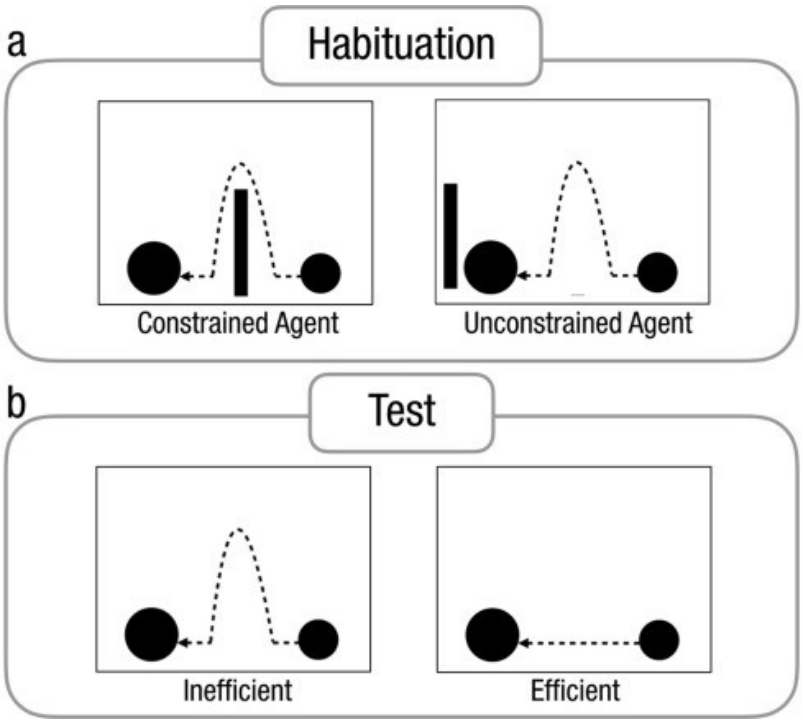

Figure 24: Interprétation téléologique de l'action d'un agent. D'abord (a), des enfants âgés de six mois sont habitués à voir le déplacement d'un petit cercle vers un grand cercle dans deux situations différentes. Ensuite (b), on leur montre une situation nouvelle. Les enfants s'attendent à chaque fois à ce que le petit cercle adopte la trajectoire la plus efficiente pour atteindre le grand cercle. (Source : Liu & Spelke, 2017).

deux situations nouvelles : dans un cas, le petit cercle se déplace comme avant, c'est-à-dire en réalisant un « saut » pour aller vers le grand cercle ; dans l'autre, il se déplace en ligne droite (voir Fig. 24). Les résultats montrent que les enfants regardent le mouvement familier plus longtemps que le nouveau, révélant ainsi leur surprise (violation des attentes). Ils s'étonnent que le petit cercle fasse le même trajet — avec et sans obstacle — au lieu d'adopter l'approche la plus directe, qui est aussi la plus efficiente. La conclusion des auteurs (Gergely & Csibra, 2003; Liu & Spelke, 2017) est que les enfants peuvent :

1. interpréter l'action d'un agent comme étant dirigée vers un but ;
2. évaluer parmi différentes actions disponibles laquelle est la plus efficiente pour atteindre le but, étant donné les contraintes de la situation ; et
3. s'attendre à ce que l'agent réalise l'action la plus efficiente à sa disposition.

L'attitude téléologique (*teleological stance*) est un système d'interprétation des actions guidé par le principe d'action rationnelle ou efficiente, emprunté à l'attitude intentionnelle de Dennett (1987) en vertu duquel les agents sont censés réaliser l'action la plus efficiente disponible, étant donné les contraintes de la situation, pour atteindre leur but. Plusieurs études ont montré que les nourrissons de 3 à 6 mois sont déjà sensibles à ce principe à l'heure d'attribuer un but à une action et que cette sensibilité pourrait également être présente chez des primates non-humains (voir Gergely, 2011). Ces résultats suggèrent que la compréhension des actions téléologiques basée sur l'évaluation de leur efficience pourrait être une adaptation cognitive relativement ancienne d'un point de vue phylogénétique.

Par ailleurs, pour qu'un objet soit reconnu comme un agent — auquel attribuer des buts — il ne faut pas seulement qu'il ait un mouvement propre (autopropulsé) mais aussi que son mouvement soit « libre » et qu'il ait la capacité d'en modifier le cours, par exemple pour s'adapter aux variations de l'environnement ou aux déplacements de sa cible (Csibra, 2008b), ce que Premack & Premack (1997) appellent la « compétence motrice ». Ces indices sont suffisants pour considérer que l'action d'un agent est dirigée vers un but, mais ils ne suffisent pas pour attribuer à cet agent des états mentaux (désirs, croyances, intentions), car une telle attribution impliquerait non seulement des capacités représentationnelles beaucoup plus sophistiquées chez l'agent, mais requerrait aussi que l'observateur ait ces mêmes capacités. L'attitude téléologique est donc moins exigeante que l'attitude intentionnelle, aussi bien en termes ontologiques que psychologiques. Dès lors, on peut envisager qu'un enfant ou qu'un animal soit capable de voir un objet comme un agent et d'interpréter son action comme un moyen en vue d'une fin, en appliquant le principe d'action efficiente, bien qu'il soit incapable de se représenter des

états mentaux intentionnels. L'attitude téléologique serait donc, selon l'hypothèse des auteurs, une faculté indépendante et peut-être antérieure à l'attitude intentionnelle (Gergely & Csibra, 2003).

L'attitude téléologique peut être définie comme un système inférentiel d'interprétation des actions qui s'appuie sur le principe d'action efficiente ou d'action efficiente, qui est sensible au contexte et qui ne dépend pas de la capacité à lire les pensées d'un agent (théorie de l'esprit), ni de l'attribution d'états mentaux (attitude intentionnelle), ni de la simulation de ses actions (par les neurones miroir). Nonobstant, cela ne veut pas dire que ces capacités ne soient pas présentes ou qu'elles ne jouent pas un rôle dans l'interprétation téléologique. En effet, on peut difficilement nier l'existence des neurones miroir, et aussitôt que l'enfant devient capable de se représenter des états mentaux, ces derniers peuvent venir enrichir sa compréhension des actions de l'agent.

Cela soulève plusieurs questions. On peut se demander, par exemple, si l'attitude téléologique ne serait pas également explicable par la théorie de la simulation en disant que l'enfant se met à la place de l'agent et qu'il considère en première personne les moyens les plus efficients pour atteindre ses fins, puis s'attend à ce que l'agent adopte ces mêmes moyens (A. I. Goldman, 2012). Depuis la même perspective, on peut aussi se demander si, au lieu de recourir au principe d'action rationnelle ou efficiente, il ne serait pas plus simple de s'appuyer sur les connaissances motrices des nourrissons, ce qui éviterait de présupposer un système inférentiel (Sinigaglia, 2008). En effet, une étude a montré que des bébés de trois mois sont capables de percevoir l'action d'un agent humain comme étant dirigée vers un but, à condition qu'ils aient eux-mêmes réalisé cette action auparavant, c'est-à-dire que la perception du but chez l'autre serait liée à leur propre expérience motrice (Sommerville, Woodward, & Needham, 2005). D'autres études suggèrent que l'attribution de buts chez les nourrissons serait limitée aux humains ou à des agents familiers. Plus généralement, il existe une controverse quant à la manière dont les enfants interprètent les actions téléologiques au cours de leur première année : s'appuient-ils sur des indices objectifs, comme le mouvement autonome, ce qui leur permettrait d'attribuer des buts à tous types d'agents et d'actions, conformément à la théorie de Gergely et Csibra, ou s'appuient-ils au contraire sur leur expérience propre, conformément aux théories de la simulation, ce qui limiterait vraisemblablement la portée de leurs attributions à des actions et des agents familiers ?

2.3. Prise de position en faveur de l'attitude téléologique

Sans vouloir préjuger de l'issue de ce débat, qui reste ouvert, nous allons prendre parti dans la suite de ce travail en faveur de la théorie des psychologues hongrois. Nos raisons sont les suivantes.

Tout d'abord, la plupart des études récentes portant sur des primates non-humains semblent confirmer cette théorie face à celle de la simulation motrice (voir Hauser & Wood, 2010). Ensuite, chez l'humain, plusieurs études ont montré que des figures géométriques et des objets inanimés, comme une simple boite, peuvent être reconnus comme des agents si leur comportement présente certains indices (Csibra, 2008b; Luo, Kaufman, & Baillargeon, 2009). De plus, les petits enfants, dès l'âge de trois mois, interprètent en termes téléologiques des actions non-familières réalisées par une main ou par un objet inanimé (un tube ou une boite) lorsqu'ils y voient certains indices, comme l'équifinalité des variations (Biro & Leslie, 2007; S. C. Johnson & Ok, 2007; Király, Jovanovic, Prinz, Aschersleben, & Gergely, 2003; Luo, 2011).

Depuis une perspective différente, des roboticiens ont montré, d'une part, que ce n'est pas l'apparence d'un artefact mais certains indices de son comportement, et en particulier sa capacité à interagir, qui nous poussent à lui attribuer des buts ou des intentions (Terada, Shamoto, & Ito, 2008; Terada, Shamoto, Mei, & Ito, 2007) et, d'autre part, qu'un robot humanoïde est traité différemment selon que son comportement est facilement prévisible ou pas (Tanaka, Cicourel, & Movellan, 2007).

Par ailleurs, deux études centrées sur les mouvements oculaires proactifs, c'est-à-dire ceux qui anticipent le mouvement d'un objet, appuient la théorie de l'attitude téléologique face à celle de la simulation par les neurones miroir (Biro, 2013; Eshuis, Coventry, & Vulchanova, 2009). Une autre étude a montré que les enfants peuvent prédire une action impossible à réaliser d'un point de vue biomécanique si elle est plus efficiente qu'une action possible, ce qui semble difficile à expliquer en termes de simulation (Southgate, Johnson, & Csibra, 2008). En outre, il semble que les neurones miroir s'activent *avant* l'observation d'une action lorsque celle-ci peut être anticipée, de sorte que leur activation ne serait pas causée par la perception visuelle d'une action en cours mais par la compréhension d'une action à venir (Southgate, Johnson, Osborne, & Csibra, 2009). Et, finalement, face à des actions familières mais partielles, il semble que les neurones miroir des enfants de 9 mois ne s'activent que lorsqu'ils peuvent en inférer le résultat (Southgate, Johnson, El Karoui, & Csibra, 2010). L'une des conclusions que tirent les partisans de l'attitude téléologique est que la fonction des neurones miroir n'est pas de simuler les actions perçues pour en permettre la compréhension, mais de les reconstruire à partir d'une interprétation téléologique préalable pour

— entre autres choses — anticiper d'éventuelles actions futures et faciliter ainsi la coopération sociale (Csibra, 2005, 2008a).

En ce qui concerne la *promiscuous teleology*, c'est-à-dire la théorie selon laquelle la pensée téléologique dérive de notre capacité à comprendre les intentions et les artefacts, les travaux de Kelemen pourraient être compatibles avec celle de Gergely et Csibra. Et effet, d'après ces derniers, lorsque les enfants deviennent capables de se représenter des états mentaux intentionnels, les attributions correspondantes (y compris les feintes et les croyances fausses) viennent compléter et enrichir leur compréhension des actions téléologiques des agents. Or, les études de Kelemen portent pour la plupart sur des enfants de plus de trois ans qui possèdent déjà une certaine capacité de représentation intentionnelle. Celles de Gergely et Csibra, en revanche, portent sur des enfants de moins de deux ans. Leurs résultats respectifs sont donc difficilement comparables puisqu'ils s'adressent à des moments différents du développement cognitif et ne posent pas les mêmes questions. Par ailleurs, les résultats des études de Kelemen sont en contradiction apparente avec ceux obtenus par Keil et d'autres auteurs qui pensent que la téléologie correspond à un module cognitif spécifique à un domaine d'objets. Dès lors, à défaut d'une explication qui permette de trancher le débat, nous retenons deux éléments importants : d'un côté, le développement d'une théorie de l'esprit — ou de capacités de compréhension et de représentation d'états mentaux intentionnels — semble être beaucoup plus précoce qu'on ne le pensait il y a quelques années ; d'un autre côté, le nombre de travaux qui plaident en faveur de l'indépendance de la téléologie vis-à-vis de la psychologie ne cesse de s'accroître.

2.4. Des modes de raisonnement indépendants

Il semble de plus en plus évident que l'esprit humain possède plusieurs systèmes innés autonomes pour interpréter le monde qui l'entoure. En particulier, nous avons un système de représentation physique des objets inanimés et de leurs interactions mécaniques. Nous avons aussi, très certainement, un système de représentation téléologique pour les agents et les actions dirigées vers un but. Ajoutons à cela un système de représentation mathématique des ensembles et de leurs relations numériques d'ordre, d'addition et de soustraction, et un autre qui nous permet de nous représenter la configuration spatiale des lieux et leurs relations géométriques (Spelke & Kinzler, 2007).

Ces systèmes ont été observées chez le nourrisson humain et chez plusieurs animaux non-humains, et pas seulement chez des primates. On peut donc vraisemblablement faire remonter leur origine phylogénétique au-delà de l'espèce humaine et de ses plus proches ancêtres.

Tout porte à croire que nous avons aussi un système psychologique pour représenter les états mentaux des agents humains et pour interpréter leurs actions ; un système dont le développement aurait lieu au cours des deux premières années de la vie de l'enfant. On sait que les nouveaux-nés ont un système pour détecter les visages (Goren, Sarty, & Wu, 1975), à l'instar de certains mammifères et invertébrés (voir Salva, Regolin, & Vallortigara, 2012), et que cette capacité a vraisemblablement une fonction sociale. On sait aussi que les petits enfants et d'autres animaux, comme les chiens et les corbeaux, prêtent une attention particulière à la direction du regard des autres et semblent leur associer des états mentaux (croyances, désirs) ce qui leur permet notamment d'interagir avec eux. Nous ne savons pas si les animaux non-humains ont aussi une théorie de l'esprit, mais il semble évident que les humains en ont une et que celle-ci se développe à un âge précoce, bien que sa nature et son origine ne soient pas encore bien connus (voir A. I. Goldman, 2012). Les enfants autistes, en revanche, seraient « aveugles à l'esprit », à savoir qu'ils auraient du mal à attribuer aux autres un esprit et les voient comme des objets matériels, ce qui expliquerait leurs difficultés ou leur manque d'intérêt pour les relations sociales (Baron-Cohen, Leslie, & Frith, 1985; Pinker, 2000). Or, l'autisme a presque certainement des causes neurologiques et génétiques. On peut donc raisonnablement penser que l'intuition psychologique qui leur fait défaut et qui les empêche d'établir des relations sociales « normales » n'est pas un produit de la culture mais de la nature.

Paul Bloom (2004) disait que les enfants sont dualistes, car leur monde se compose de deux types d'entités : les objets inanimés, dont le comportement obéit à des principes physiques, et les esprits, mus par des émotions et des intentions. Cependant, l'une des conclusions auxquelles nous sommes arrivés dans la section précédente est qu'il existe pour eux un troisième type d'entités, celle des agents qui sont mus par des fins et dont le comportement obéit à un principe de rationalité, mais auxquels les enfants n'attribuent pas nécessairement des états mentaux. Nous avons vu que la téléologie naïve n'est ni un produit dérivé de la psychologie ni une forme immature de la causalité. Au contraire, c'est un mode d'interprétation de la réalité à part entière, aussi primordial que la pensée physique et peut-être plus fondamental que la pensée intentionnelle.

3. Aux origines du raisonnement téléofonctionnel

3.1. Le raisonnement téléofonctionnel est-il lié à l'attitude téléologique ?

Dans quelle mesure les conclusions précédentes nous aident-elles à comprendre la nature et l'origine du biais cognitif téléofonctionnel ? On peut formuler l'hypothèse que ce biais dérive ou dépend du système de représentation téléologique et non pas, comme l'affirme Kelemen, du système de représentation intentionnelle. Selon cette hypothèse, nous adoptons l'attitude téléologique vis-à-vis des artefacts comme nous le faisons pour les actions. En effet, chez les animaux non-humains, ce biais se manifeste principalement — et peut-être uniquement — à travers l'usage et la fabrication d'outils. Or, un outil est un moyen d'une fin ; et de même qu'il y a des objets qui font obstacle à la réalisation d'une action, il y en a d'autres qui la facilitent ou qui la rendent possible. On peut donc penser que les outils s'inscrivent dans le cadre de l'action téléologique et qu'ils relèvent du mode de représentation correspondante[153].

Lorsque les mésanges bleues enlèvent la capsule des bouteilles de lait pour en manger la crème, elles agissent sur leur environnement physique pour se défaire d'un obstacle et arriver à leurs fins. Lorsque les singes capucins cassent des noix de palme avec un marteau et une enclume, ils font la même chose, à ceci près que leur action inclut la manipulation de deux outils de pierre. Certes, l'usage de l'outil change beaucoup de choses, mais, dans les deux cas, il s'agit d'actions dirigées vers une fin, et c'est sans doute en tant que telles que leurs congénères les voient et les interprètent, notamment lorsqu'ils apprennent par imitation ou par émulation à les reproduire. Depuis cette perspective, l'outil serait un composant de l'action et serait donc soumis au principe de rationalité que lui attribue la théorie de l'attitude téléologique. On sait par exemple que les animaux sont capables de distinguer les propriétés fonctionnelles pertinentes d'un objet pour la réalisation d'une tâche, c'est-à-dire celles qui font que l'objet soit un moyen efficient de celle-ci (Clair & Rutz, 2013; Hauser, 2002; Herrmann, Wobber, & Call, 2008; Ruiz & Santos, 2013). Une étude a également montré que les jeunes chimpanzés qui observent un adulte réaliser une tâche avec un outil sont capables de comprendre les relations causales entre l'outil et ses conséquences et qu'ils ne reproduisent que les actions pertinentes, c'est-à-dire efficientes, pour l'obtention du résultat (Horner & Whiten, 2005).

153 À propos du caractère intentionnel ou pas des outils animaux, voir (Shumaker, Walkup, & Beck, 2011; St Amant & Horton, 2008). À propos des artefacts animaux en général, voir Gould (2007).

Cela nous conduit à penser que des animaux comme les chimpanzés « comprennent » les propriétés fonctionnelles des outils d'une façon qui est compatible avec la définition du concept de fonction proposée au CHAP. VII. L'étude citée ci-dessus présente une situation expérimentale où de la nourriture est placée dans un tube à l'intérieur d'une boite et n'est accessible qu'au travers d'un trou ; dans ce contexte, les chimpanzés doivent manipuler un outil d'une certaine façon pour extraire la nourriture de l'intérieur de la boite. Si l'on considère cette tâche comme un système d'actions dirigé vers une fin, alors la fonction de l'outil dans ce système est d'accéder à la nourriture pour pouvoir l'extraire, selon l'énoncé de la définition :

> La/une fonction d'un *token* de type *X* dans un système de type *S* et dans le contexte *C* est $f_{(X)}$.

Les chimpanzés savent que l'outil doit avoir certaines caractéristiques physiques pertinentes (longueur, diamètre, rigidité, etc.) et qu'il doit être employé de manière efficiente. Autrement dit, ils savent que l'extraction est une conséquence causale de l'usage de l'outil et de ses propriétés physiques, selon la première condition de la définition :

> (i) $f_{(X)}$ est une conséquence de *X*.

Il comprennent aussi que les actions de l'adulte ne sont pas arbitraires ni aléatoires, mais qu'elles visent à obtenir un résultat, et que pour y arriver il faut les réaliser dans un certain ordre et d'une certaine manière ; c'est-à-dire que la série d'actions de l'adulte constitue un système dirigé vers une fin — manger la nourriture —, et que l'utilisation de l'outil pour l'extraire de la boite contribue à cette fin :

> (ii) $f_{(X)}$ contribue à une fin *F* de *S*.

Et dans la mesure où ils ne reproduisent pas toutes les actions observées mais seulement celles qui sont pertinentes pour l'obtention du résultat souhaité, ils se montrent capables de distinguer les actions qui appartiennent effectivement à la tâche, c'est-à-dire au système, de celles qui sont superflues, inutiles ou inefficientes et qui par conséquent n'y appartiennent pas :

> (iii) $f_{(X)}$ appartient à *S*.

On retrouve de cette façon les trois conditions de la définition : la relation causale entre l'élément fonctionnel et ses effets, la relation moyens-fin, c'est-à-dire la contribution de la fonction à une fin du système, et la relation partie-tout, c'est-à-dire l'appartenance de la fonction au système.

Toutefois, le mode de représentation téléofonctionnelle de l'enfant humain pourrait ne pas s'ajuster à cette définition ni dépendre seulement de l'attitude téléologique. En effet, il existe au moins quatre différences

essentielles entre le comportement des animaux et celui des enfants en ce qui concerne l'apprentissage et l'utilisation des outils.

3.2. Quatre différences entre les humains et les autres animaux

La première différence est celle que révèle l'étude précédente : contrairement aux chimpanzés, les enfants imitent fidèlement toutes les actions réalisées par le démonstrateur sans tenir compte de leur efficience, au lieu de ne reproduire que celles qui contribuent effectivement au résultat et alors qu'ils ont les connaissances causales suffisantes pour distinguer les unes des autres (Horner & Whiten, 2005). Une étude postérieure visant à répliquer et élargir ces résultats avec des enfants de trois et cinq ans dans des conditions expérimentales mieux contrôlées a montré que les humains ne deviennent pas plus sélectifs avec l'âge, malgré leur meilleure compréhension des relations causales, mais tendent au contraire à imiter davantage la séquence d'actions observée en y incluant celles qui sont manifestement superflues (McGuigan, Whiten, Flynn, & Horner, 2007). Et il semble que cette tendance à la sur-imitation s'accentue encore chez l'adulte, lequel va copier à la perfection tous les gestes du démonstrateur sans distinguer ceux qui sont pertinents de ceux qui ne le sont pas, et qui va par conséquent réaliser la même tâche de manière moins efficiente qu'un enfant de trois ans (McGuigan, Makinson, & Whiten, 2011).

La seconde différence est celle que signalent Hernik et Csibra (2009) quant à la manière de concevoir un outil lorsqu'il n'est pas utilisé. Les humains ont la capacité de fabriquer et de stocker des outils pour un usage ultérieur, ce qui a rarement été observé chez l'animal[154], et ils ont aussi la faculté de penser aux usages potentiels d'un objet en dehors de toute nécessité immédiate. En revanche, il ne semble pas que les autres animaux, même ceux qui savent utiliser et fabriquer des outils, conçoivent un objet en termes fonctionnels lorsqu'ils ne l'utilisent pas. Et bien qu'ils soient capables de raisonner en termes de moyens potentiels pour atteindre un but donné, c'est-à-dire pour satisfaire un besoin ou un désir immédiat, ils n'envisagent pas des buts potentiels face à un objet donné. Or, c'est là précisément le défi auquel sont confrontés les enfants lorsqu'ils entrent dans le monde des artefacts définis culturellement, car ces derniers sont des moyens de fins non spécifiées. Un couteau, par

154 Un chimpanzé du zoo de Stockholm a été observé en train de ramasser tranquillement des cailloux, les façonner et les stocker à des endroits stratégiques et à l'abri des regards pour les lancer quelques heures après contre les visiteurs (Osvath, 2009; Osvath & Karvonen, 2012). Des corbeaux de Nouvelle Calédonie, quant à eux, prennent soin de ne pas perdre leurs outils — des brindilles servant à extraire des insectes — pour les réutiliser plus tard (Klump, Wal, Clair, & Rutz, 2015).

exemple, sert à couper, mais le fait de couper n'est pas une fin en soi ; ce n'est que le moyen de multiples fins possibles : couper un morceau de viande pour le manger, du cuir pour en faire un vêtement, une branche pour en faire un outil, etc.

Une troisième différence, liée à la précédente, est que les singes n'assignent pas à un objet un usage déterminée. Pour répondre à un besoin, ils utilisent comme outil n'importe quel objet ayant des propriétés fonctionnelles adéquates, le même objet pouvant ensuite être utilisé pour autre chose. Ils semblent donc traiter les outils comme de simples moyens en vue d'une fin. Au contraire, les humains ont tendance à penser que chaque outil a une fonction propre. Ainsi, nous évitons d'utiliser un objet pour une fonction qui ne lui correspond pas, bien qu'il ait les propriétés fonctionnelles adéquates. Par exemple, nous avons des fourchettes pour le poisson, pour la viande, pour le dessert, etc., et les règles de bienséance nous interdisent formellement de confondre les unes avec les autres. Dès l'âge de six à huit mois, les enfants sont capables de choisir les moyens appropriés pour atteindre leurs fins (Willatts, 1999). Dès l'âge de deux ans, ils forment des catégories stables où chaque artefact est *pour* un usage donné, et il leur suffit d'une seule démonstration pour comprendre à quelle catégorie appartient un artefact nouveau (Casler & Kelemen, 2005, 2007).

La quatrième différence a trait au processus d'apprentissage. Chez les primates, les petits observent les gestes réalisés par un autre membre du groupe et essayent de les reproduire. Il n'y a pas d'enseignement à proprement parler dans la mesure où l'« enseignant » ne fait rien de particulier à destination de l'apprenant, et en particulier il ne semble pas qu'il altère la séquence motrice de réalisation de la tâche. L'apprentissage des vocalisations par les jeunes ouistitis est le seul exemple documenté jusqu'à maintenant d'enseignement animal et pourrait être un précurseur de l'apprentissage du langage humain (Margoliash & Tchernichovski, 2015; Takahashi et al., 2015). Chez ce dernier, l'enseignant adapte ses gestes, il les ralentit, les répète, les décompose, et il emploie une série d'indices pour que le petit comprenne qu'on veut lui enseigner quelque chose : contact visuel réciproque, ton de la voix, etc. (voir Gergely, 2011) Les humains et les chiens sont de fait les seules espèces connues capables de comprendre certaines formes de communication référentielle comme pointer du doigt pour signaler une chose que l'on veut montrer (Hare & Woods, 2013).

3.3. Dimensions intentionnelle et sociale des artefacts et catégorisation fonctionnelle

Les hypothèses avancées pour expliquer ces différences sont toutes liées au caractère fortement intentionnel et social de la culture matérielle chez l'humain. Si nous copions les gestes du démonstrateur avec l'outil,

c'est peut-être parce que nous le considérons comme un expert qui sait ce qu'il fait et qui les réalise volontairement ; ou parce qu'il veut nous les montrer pour une raison ou pour une autre ; ou nous le faisons simplement pour interagir socialement avec lui et lui faire plaisir ; ou encore, et c'est l'hypothèse privilégiée par les auteurs de l'étude, parce que nous aurions une disposition croissante avec l'âge à imiter automatiquement les actions d'autrui, quelles qu'elles soient, car cela aurait une forte valeur adaptative en termes d'apprentissage, d'acculturation, de renforcement du lien social, etc. (McGuigan et al., 2011).

Lorsqu'ils doivent utiliser un outil pour réaliser une tâche, le choix des humains pour un objet ou pour un autre ne dépend pas seulement de son adéquation fonctionnelle, c'est-à-dire des caractéristiques physiques qui en font ou pas un bon moyen pour cette fin, mais il est guidé aussi et surtout par l'information sociale dont ils disposent : s'ils ont eu l'occasion de voir quelqu'un réaliser intentionnellement cette même tâche avec un objet donné, ils préfèrent utiliser le même objet, bien que des alternatives également fonctionnelles soient plus facilement accessibles, et ils évitent de l'utiliser pour des tâches différentes (Casler & Kelemen, 2005). En cela, les adultes et les enfants de deux ans se comportent de la même façon. C'est le signe d'une catégorisation fondée sur l'usage intentionnel ou social des artefacts plutôt que sur leurs dispositions causales.

La constitution des catégories d'artefacts est une question controversée, mais les études menées par Deborah Kemler Nelson et d'autres auteurs tendent à montrer que dès l'âge de deux ans les enfants nomment les artefacts d'après leur fonction (leur usage) et pas seulement d'après leur apparence générale ou leur forme (Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson, Herron, & Holt, 2003; Kemler Nelson, Russell, Duke, & Jones, 2000), et quand ils s'interrogent sur un nouvel artefact, l'information qui les intéresse en premier lieu est la fonction de l'objet : « ce que ça fait », « comment ça marche », etc. (Greif et al., 2006; Kemler Nelson et al., 2004; Kemler Nelson, O'Neil, & Asher, 2008). Par ailleurs, dès l'âge de trois ou quatre ans, ils semblent être capables d'inférer la fonction prétendue d'un artefact (ce pour quoi il est fait) à partir de ses caractéristiques physiques y compris lorsqu'il est cassé (Asher & Kemler Nelson, 2008; Kemler Nelson, Herron, & Morris, 2002). Ces résultats sont cohérents avec ceux obtenus par Krista Casler et Deborah Kelemen (2005, 2007). Ils indiquent que la catégorisation des artefacts s'appuie effectivement sur la connaissance de leur fonction et que celle-ci est acquise avant tout par le biais d'interactions sociales. Ainsi, nous présupposons que tous les artefacts sont *pour* quelque chose et nous prêtons une attention particulière à l'usage (intentionnel) qu'en font les autres ou à ce qu'ils nous en disent pour les classer dans la catégorie fonctionnelle qui leur correspond (socialement).

Cela étant, la compréhension intuitive du concept de fonction évolue vraisemblablement avec l'âge. En effet, il semble que les adultes adoptent vis-à-vis des artefacts ce que Daniel Dennett (1971, 1987, 1990) appelle le *design stance*, c'est-à-dire une attitude ou une interprétation selon laquelle les choses ont été créées à dessein et avec un plan d'ensemble, une structure interne qui, bien qu'elle ne nous soit pas connue, nous permet néanmoins de réaliser certaines prédictions quant à son comportement. Ainsi, lorsque quelqu'un programme l'alarme d'un réveil-matin pour 8:00, il s'attend à ce qu'il sonne le lendemain à l'heure indiquée, et il n'est pas nécessaire pour cela d'ouvrir l'appareil et d'en analyser les mécanismes internes. Il lui suffit de supposer qu'il a été créé dans ce but et que, sauf défaut de fabrication, panne mécanique ou erreur de manipulation, il fera *ce pour quoi il a été conçu*. C'est là le sens du concept de fonction depuis le *design stance*.

Il y a cependant une ambiguïté chez Dennett concernant la relation entre le *design stance* et l'*intentional stance*, car en parlant des objets biologiques, il semble affirmer et nier simultanément l'intentionnalité de « Mère Nature » (voir CHAP. XII, SECT. 5, p.387), et bien qu'il affirme que l'attitude du *design* est plus fondamentale que l'attitude intentionnelle, on peut montrer de manière assez convaincante que la première implique au contraire la seconde (Ratcliffe, 2001). Pieter Vermaas et ses collègues(2013) proposent donc de lever l'ambiguïté en distinguant deux attitudes différentes, faible et forte, nommées respectivement *teleological design stance* et *intentional designer stance*. La première implique qu'une entité a des parties avec des fonctions relatives à une fin (*purpose*). La seconde implique que l'entité a été conçue par un agent intentionnel pour avoir ces fonctions et cette fin.

D'après certains auteurs, la conception selon laquelle la fonction d'un artefact est ce pour quoi il a été conçu n'apparaît pas avant l'âge de quatre à six ans, au terme d'un processus de changement conceptuel traversant plusieurs étapes encore mal connues (Casler & Kelemen, 2005; German & Johnson, 2002; Kelemen & Carey, 2007). Dans la première, les enfants pourraient aborder les artefacts à partir de leur compréhension des actions téléologiques comme de simples moyens pour arriver à une fin, à la manière de certains primates. Dans la deuxième, ils pourraient associer chaque objet à un usage déterminé et, à mesure que leur compréhension des actions intentionnelles se développe, ils distingueraient les usages accidentels et intentionnels. Ensuite, vers l'âge de deux ans, ils commenceraient à former des catégories stables à partir de l'idée que les objets ont des fonctions qui leur sont intrinsèques — de sorte que tout le monde aurait besoin des mêmes objets pour réaliser la même fonction — et ils s'appuieraient sur l'observation de l'usage qu'en font les autres pour découvrir ces fonctions. Mais à ce stade la fonction des artefacts n'est pas encore liée à leur histoire. Ainsi, par exemple, si quelqu'un utilise un objet pour boire, alors cet objet est classé dans la catégorie des verres,

bien qu'il ait été créé pour être un vase. C'est vers trois ans que les enfants se rendraient compte que la fonction propre d'un artefact est invariable et qu'elle ne dépend pas de l'usage qu'on en fait. Si un objet est un vase, alors il ne peut pas devenir un verre même si quelqu'un l'utilise de cette façon. De là viendrait le droit de baptême accordé au créateur d'artefacts, car c'est lui qui fixe la fonction d'un objet au moment où il le crée, celle-ci ne pouvant plus dès lors être modifiée. Finalement, ils identifieraient la fonction non plus seulement à l'intention originale du créateur mais au *design* original de l'objet, c'est-à-dire à ce pour quoi il a été conçu. En somme, le biais téléofonctionnel s'appuierait à la fois sur les systèmes de représentation physique, téléologique et intentionnel, et sur un processus de construction théorique général qui pousse l'enfant à essentialiser les artefacts et à les catégoriser selon leur origine.

3.4. La théorie de la pédagogie naturelle

Les travaux de Csibra et Gergely (2006, 2009, 2011) laissent entrevoir une autre explication de l'origine du biais téléofonctionnel. Une explication qui s'appuie sur deux systèmes cognitifs innés et initialement indépendants ayant évolué comme des adaptations séparées pour la représentation des deux types d'agents qui peuplent notre environnement socio-culturel humain, à savoir les « agents instrumentaux » ou téléologiques, d'un côté, et les « agents communicatifs », de l'autre (voir Gergely, 2011). Le premier des deux systèmes est l'attitude téléologique dont nous avons déjà parlé, et les agents instrumentaux sont ceux dont le comportement peut être interprété à la lumière du principe d'action rationnelle ou efficiente comme un moyen en vue d'une fin. L'autre système est ce qu'ils appellent la « pédagogie naturelle », une adaptation cognitive spécialisée pour permettre l'apprentissage social rapide et efficient de connaissances culturelles cognitivement opaques qui seraient difficiles à acquérir autrement. C'est un système de communication propre à l'espèce humaine qui s'appuie sur des indices ostensifs comme le contact visuel, le fait de sourciller et de se diriger à l'autre en langage enfantin (*motheresse*), suivis de gestes référentiels déictiques, comme le fait de porter son regard vers ce qu'on veut montrer ou de le pointer du doigt, destinés à aider l'apprenant à identifier le référent dont on veut dire ou montrer quelque chose. Par exemple, pour apprendre à l'enfant à manger sa soupe à la cuillère, les parents peuvent d'abord attirer son attention sur l'objet en l'agitant en l'air ou en le pointant du doigt tout en en maintenant le contact visuel et en lui parlant sur un ton particulier, puis ils peuvent plonger la cuillère dans la soupe et la porter à la bouche en ralentissant, en répétant et en exagérant leurs gestes tout en continuant à parler en langage enfantin. Ce type de comportement est ce qui caractérise un agent communicatif.

D'après les auteurs de cette théorie, la pédagogie naturelle ne requiert ni un langage ni une théorie de l'esprit avancée, et ils formulent l'hypothèse selon laquelle son évolution pourrait avoir été rendue nécessaire par le développement de l'usage et de la fabrication d'outils chez nos ancêtres, et en particulier par l'utilisation d'outils pour fabriquer d'autres outils, ce qui implique un raisonnement téléologique récursif et des chaines ou des hiérarchies de moyens et de fins qui rendent les propriétés fonctionnelles opaques et difficiles à inférer pour un observateur non-informé (Csibra & Gergely, 2006, 2011).

L'une des caractéristiques qui distinguent la pédagogie naturelle d'autres systèmes de communication animale est sa capacité à transmettre des connaissances générales à propos de genres, comme l'utilisation ou les propriétés d'un *type* d'artefact, et pas seulement des faits épisodiques concernant un objet donné ou une situation concrète (Csibra & Gergely, 2009). Le fait de se diriger à l'enfant avec des indices ostensifs induit chez lui un biais interprétatif qui l'amène à voir l'information transmise comme ayant une valeur générale. Ainsi, par exemple, lorsque la démonstration d'une tâche s'accompagne de signes ostensifs référentiels, les enfants de 14 mois ont tendance à apprendre par imitation les actions cognitivement opaques de cette tâche — ce qui leur donne les moyens de faire certaines choses avant même de comprendre le rôle causal que jouent ces moyens dans l'obtention de la fin qu'ils désirent — tandis que la même tâche observée dans un contexte non communicatif n'entraine pas l'imitation des actions opaques (voir Gergely, 2011, p. 82). En ce qui concerne l'interprétation des objets, les signes ostensifs référentiels semblent inhiber le traitement de l'information spatiale et favorisent la perception de caractéristiques physiques (comme la forme) qui ont des chances d'être pertinentes pour l'identification de propriétés généralisables aux objets du même type. De plus, la catégorisation des artefacts peut être induite par la démonstration non verbale de leur usage fonctionnel, dès l'âge de 10 mois, lorsque cette démonstration est précédée de signaux ostensifs (Futó, Téglás, Csibra, & Gergely, 2010).

3.5. Conclusions concernant les artefacts et le vivant

Les différentes théories que nous venons d'exposer brièvement quant à l'origine du biais téléofonctionnel sont sans doute complémentaires plutôt que contradictoires. Prises conjointement, on peut en tirer les conclusions suivantes. La première est que les enfants humains sont capables de catégoriser les artefacts selon leur fonction — c'est-à-dire tout d'abord selon leur usage dans le cadre d'une relation moyens-fin — au plus tard à l'âge de 2 ans et peut-être déjà vers l'âge de 10 mois, c'est-à-dire de façon assez précoce étant donné que les autres primates n'en sont apparemment jamais capables. Et si cette capacité n'est pas elle-même

innée, elle se développe néanmoins sur la base de mécanismes qui eux le sont.

La seconde est que les catégories fonctionnelles ont une dimension sociale très marquée qui se manifeste non seulement à travers l'apprentissage, puisque leur connaissance est transmise par le biais d'interactions sociales et pas uniquement par l'observation de l'action d'autres membres du groupe, mais aussi par le fait qu'elles appartiennent à une culture commune. En effet, si les fourchettes à poisson sont pour manger le poisson — alors que n'importe quelle autre fourchette pourrait faire l'affaire et qu'une fourchette à poisson peut servir à manger n'importe quoi d'autre[155] —, c'est parce que leur fonction repose sur une convention culturelle : parmi toutes les choses auxquelles elles peuvent servir, leur fonction est celle que la communauté reconnaît comme telle. Cela implique aussi que les concepts fonctionnels, comme celui de fourchette, sont partagés socialement, de sorte que si quelqu'un veut en fabriquer un nouveau modèle, il doit s'ajuster au concept préexistant ou négocier son inclusion.

La troisième conclusion est que les fonctions ont aussi une dimension intentionnelle très importante qui se manifeste aussitôt que la théorie de l'esprit est en place. Elle se manifeste d'abord dans l'apprentissage, par exemple dans la distinction entre les usages accidentels et intentionnels d'un artefact, et ensuite dans la détermination des catégories fonctionnelles, puisqu'au cours du développement cognitif, avant l'apparition du *design stance*, la fonction d'un objet semble être associée d'abord à l'usage intentionnel qui en est fait puis aux intentions originales de son créateur.

La quatrième conclusion que l'on peut tirer est qu'il existe en principe un rapport causal entre la fonction d'un objet et ses caractéristiques physiques ; un rapport dicté par l'adéquation entre les moyens et les fins. Nous avons vu que les chimpanzés et les jeunes enfants comprennent ce rapport et qu'ils reconnaissent les propriétés fonctionnelles des outils, c'est-à-dire celles qui permettent d'obtenir de manière efficiente le résultat visé. Nous avons vu aussi que vers 3-4 ans, les enfants sont capables d'inférer la fonction prétendue d'un artefact à partir de ses caractéristiques physiques. Et comme les artefacts humains sont généralement

155 Nous ne voulons pas dire pour autant que les formes et les tailles des différentes fourchettes soient arbitraires, car elles ont sans doute été conçues en fonction du type d'aliment auquel elles sont destinées. Par exemple, si les dents de la fourchette à poisson sont plus larges, c'est peut-être parce que celui-ci a une chair trop tendre pour être piquée comme celle de la viande et qu'il faut recueillir comme on le ferait avec une pelle. Et si elles n'ont que trois dents au lieu de quatre, c'est peut-être pour rappeler le trident de Neptune. Une autre théorie veut que l'on utilise des couverts spéciaux pour le poisson car les propriétés chimiques de celui-ci altéraient l'argent des couverts habituels.

spécialisés et optimisés pour la réalisation la plus efficiente possible d'une tâche particulière, on peut souvent en inférer la fonction — lorsque celle-ci ne nous a pas été transmise socialement — à partir de l'observation de leurs propriétés. C'est ce que font habituellement les archéologues face à des objets aussi complexes que le mécanisme d'Anticythère et aussi simples que les outils lithiques de la préhistoire. Mais ce qui distingue le raisonnement fonctionnel humain de celui des autres animaux est que ces derniers s'appuient seulement sur la pensée téléologique, tandis que nous disposons aussi de la pensée intentionnelle et de la pédagogie naturelle qui nous permettent de former des catégories fonctionnelles stables et socialement partagées.

La cinquième conclusion que nous voulons retenir est que les enfants conçoivent rapidement les fonctions comme des propriétés intrinsèques des artefacts ; une conception qui perdure à l'âge adulte alors même que nous reconnaissons par ailleurs le caractère intentionnel et conventionnel, et par conséquent extrinsèque, des attributions fonctionnelles. Il est possible aussi, suivant Bloom (2004), Kelemen & Carey (2007) et d'autres auteurs, que nous soyons essentialistes vis-à-vis des artefacts, de même que nous le sommes vis-à-vis des genres naturels. Cela implique notamment que les groupements catégoriels ne s'appuient pas sur des caractéristiques superficielles et observables comme la forme, mais sur une constitution interne ou une « essence » qui, dans le cas des artefacts, ne serait autre que la fonction prétendue originale, c'est-à-dire ce pour quoi ils ont été créés[156]. Selon cette hypothèse, au-delà des aspects physique, intentionnel et social que nous venons de mentionner, la fonction d'un artefact serait directement liée à son origine causale.

La dernière conclusion est négative. Nous ignorons si l'origine du raisonnement téléofonctionnel dans le domaine biologique est liée à celle de la pensée instrumentale (utilitaire), si l'une dépend de l'autre ou si elles sont indépendantes. Certains auteurs, comme Deborah Kelemen, soutiennent que les outils mentaux qui nous permettent de penser les artefacts, une fois disponibles, sont généralisés à d'autres domaines d'objets comme les organismes vivants. Cela nous semble plausible dans la mesure où nous ne voyons pas quel avantage évolutif procurerait le fait d'attribuer des fonctions aux parties des animaux et des plantes considérés en eux-mêmes et pas du point de vue de leur usage instrumental (alimentaire, médicinal, vestimentaire). Cependant, nous manquons de preuves pour étayer cette hypothèse. De plus, les dimensions intentionnelle et sociale des fonctions techniques ne plaident pas en faveur d'une telle généralisation, à moins de supposer par exemple que les fonctions biologiques ne se rapportent elles-aussi aux intentions d'un créateur non-humain. En ce sens, certains auteurs affirment que la pensée téléologique

156 Pour un aperçu des débats sur l'ontologie des artefacts depuis la philosophie de la technologie, voir Vega Encabo (2009).

serait une réponse aux menaces du monde naturel et que son hyperactivité serait ce qui nous pousse à détecter des agents surnaturels, un comportement présentant beaucoup d'avantages et presque pas d'inconvénients (Atran & Norenzayan, 2004; Barrett, 2000; Girotto, Pievani, & Vallortigara, 2014). D'un autre côté, nous savons que les jeunes enfants distinguent sans difficulté les objets biologiques des autres objets naturels et artificiels et qu'ils ont des attentes spécifiques vis-à-vis des uns et des autres, à tel point que de nombreux auteurs défendent l'existence d'une biologie naïve autonome. Depuis cette perspective, la pensée téléofonctionnelle proprement biologique pourrait ne pas dépendre de celle des artefacts. Quoi qu'il en soit, et en attendant que de nouvelles données fassent avancer le débat, nous exprimons une certaine réserve quant à l'extrapolation des résultats de la psychologie des artefacts à celle des êtres vivants, mais cela ne doit pas nous interdire pour autant de chercher à mieux comprendre les fonctions biologiques à la lumière des connaissances acquises sur les fonctions techniques.

4. Le débat sur les fonctions et la psychologie cognitive

Dans quelle mesure les résultats obtenus en psychologie cognitive sont-ils pertinents pour le débat philosophique qui nous occupe ? D'une part, nous avons vu que certains auteurs situent leur définition des fonctions dans le cadre d'une analyse du concept tel que l'emploient les biologistes. Pourtant, les définitions formulées dans ce cadre ne viennent pas accompagnées d'une étude empirique qui permettrait de les justifier. Tous les biologistes emploient-ils le concept conformément à la définition proposée ou seulement une majorité d'entre eux ? Y a-t-il parmi les biologistes un seul usage du concept ou plusieurs ? Le cas échéant, les différentes utilisations du concept varient-elles selon la discipline pratiquée (microbiologie, génétique des populations, écologie, physiologie, taxinomie, etc.) ou selon d'autres facteurs ? À défaut de ce type d'études, on peut se reporter à celles qui portent sur les enfants et les adultes non spécialistes et comparer les définitions philosophiques avec les intuitions courantes.

D'autre part, nous avons vu que le débat sur les fonctions consiste en grande partie à défendre et à rejeter des définitions à grands coups d'exemples et de contre-exemples. Or, ces derniers ne reposent pas sur des faits empiriques, mais sur des intuitions. Pourquoi le trou dans le tuyau de gaz toxique est-il un contre-exemple de la définition de Wright ? Il semble que nous soyons tous d'accord pour reconnaître que le trou n'est pas là pour laisser s'échapper le gaz, qu'il n'a pas cette fonction, mais comment peut-on le savoir si ce n'est de manière intuitive ? D'un côté, certaines affirmations semblent tellement évidentes qu'elles se passent de

justification. De l'autre, les auteurs semblent incapables de s'accorder sur des points essentiels. Ainsi, pour les partisans de l'approche étiologique, il est évident qu'un cœur malformé, incapable de pomper le sang, possède la fonction qui correspond à son type, tandis que pour les partisans de l'approche dispositionnelle, cela n'a rien d'évident, bien au contraire.

Le débat entre les uns et les autres ne se résume pas à un conflit d'intuitions, mais celles-ci jouent un rôle important dans le débat. Depuis cette perspective, les études de psychologie cognitive peuvent nous aider à comprendre certaines choses. Par exemple, pourquoi les conceptions étiologique et systémique sont les stratégies de naturalisation des fonctions les plus populaires.

Tania Lombrozo et Susan Carey (2006) ont en effet réalisé une série d'études sur des étudiants de Harvard qui montrent que les explications fonctionnelles sont jugées acceptables lorsque la fonction invoquée est causalement responsable de la présence du trait fonctionnel, conformément à la conception étiologique de Larry Wright[157]. Leurs résultats confirment aussi une idée que nous avons défendue tout au long de ce travail, à savoir que les explications physiques, téléologiques et intentionnelles ne sont pas concurrentes mais complémentaires[158]. En revanche, les auteures ne parviennent pas à montrer que l'acceptabilité des explications téléologiques chez l'adulte soit indépendante du domaine, ce qui viendrait contredire les théories de Atran (1995) et de Keil (1992, 1995), car tous leurs exemples impliquent des objets biologiques ou des artefacts.

Grant Gutheil et ses collègues (2004) ont par ailleurs étudié la manière dont les enfants et les adultes catégorisent les artefacts en tenant compte de leur histoire ou de leurs propriétés actuelles. Bien qu'elles ne permettent pas de comparer directement les théories étiologique et dispositionnelle de Wright et de Cummins, ces études permettent néanmoins de confronter certaines de leurs intuitions avec celles des non spécialistes. Dit-on d'un objet que c'est un « couteau » parce qu'il a une forme typique de couteau et parce qu'il peut être utilisé comme couteau typique, ou bien la catégorie dans laquelle on le classe dépend-elle plutôt

157 Une autre condition est requise : il faut que le processus causal en question soit suffisamment général pour être prédictible. Nous reviendrons sur ce point à la section 5, p. 355.

158 Les chercheurs ont présenté aux étudiants plusieurs histoires causales différentes (intentionnelle, non intentionnelle, accidentelle) pour un même trait (biologique ou non) et leur ont demandé d'évaluer pour chacun d'elles l'acceptabilité de plusieurs explications (intentionnelle, téléologique, mécanique). Les résultats indiquent que les explications mécaniques sont acceptables dans tous les cas par la grande majorité des sujets, que les explications intentionnelles le sont lorsque l'histoire causale implique une intention, et que les explications téléologiques ne sont jugées inacceptables par la plupart des sujets que dans le scénario accidentel, c'est-à-dire lorsque la fonction n'intervient pas dans l'étiologie du trait.

de son origine et plus généralement de son histoire ? Pour confronter les deux hypothèses, les chercheurs se sont demandés si un couteau modifié ou cassé n'ayant plus la forme ni les capacités correspondantes continuerait d'être considéré comme tel. Si la catégorisation dépend de l'histoire, alors la réponse doit être positive. Au contraire, si elle dépend de la forme ou des propriétés actuelles, alors un couteau qui ne peut plus couper cesse d'être un couteau. Leurs résultats indiquent que les enfants et les adultes tiennent compte à la fois de l'histoire et des propriétés, mais pas de la même façon. Lorsque l'état actuel d'un objet est en conflit direct avec son histoire, les enfants d'âge préscolaire privilégient l'état actuel, les adultes privilégient l'histoire et les enfants de 6 à 9 ans sont partagés entre les deux. C'est-à-dire qu'au cours du développement cognitif, l'histoire d'un objet devient de plus en plus importante pour déterminer la catégorie à laquelle il appartient. Cependant, plus un objet est sévèrement détruit et moins les adultes ont tendance à l'accepter comme membre de sa catégorie. L'explication la plus probable de ces résultats, d'après les auteurs, est que les enfants et les adultes disposent de plusieurs théories ou biais explicatifs (comme la biologie naïve) pour interpréter les phénomènes d'un domaine donné (êtres vivants, artefacts) et les emploient dans des proportions différentes selon leur âge (voir aussi Gutheil et al., 1998).

Une autre explication possible de ces résultats dérive de ce que nous avons vu précédemment, à savoir que les enfants commencent à catégoriser les artefacts en tenant compte de leur histoire vers l'âge de 4 à 6 ans et que leur conception des fonctions techniques dépend d'abord de l'origine intentionnelle et ensuite du *design* originel, c'est-à-dire de ce pour quoi les artefacts ont été conçus (SECT. 3.3). Cette conception serait la base du *design stance* qui caractérise d'après Dennett (1987) et d'autres auteurs (German & Johnson, 2002; Kelemen & Carey, 2007) la pensée téléofonctionnelle de l'adulte. Une formulation particulière de cette idée, due à Paul Bloom (1996), dit qu'un artefact appartient à une catégorie donnée lorsqu'il a été créé avec succès avec l'intention d'y appartenir. Cela implique que l'étiologie d'un artefact est plus importante que ses propriétés physiques et ses fonctions ou capacités actuelles. D'un côté, un objet peut donc être une chaise ou une horloge même s'il ne ressemble pas aux autres membres de sa catégorie ; et il continue d'en faire partie même lorsqu'il est cassé et ne peut plus remplir la fonction qui lui correspond. D'un autre côté, deux objets identiques quant à leur apparence, structure et fonction n'appartiennent pas à la même catégorie s'ils n'ont pas l'histoire qu'il faut.

Il est facile de remarquer des similitudes entre cette conception des artefacts et l'approche étiologique des fonctions biologiques. Celles-ci ne seraient pas étonnantes s'il était vrai que la pensée téléofonctionnelle dans le domaine du vivant est directement liée à notre compréhension des artefacts. Quoi qu'il en soit, la psychologie cognitive semble confirmer le

caractère intuitif de la théorie de Wright et dans une moindre mesure de celle de Cummins, ce qui en retour semble correspondre aux degrés d'acceptation respectifs de ces théories parmi les biologistes.

Il serait toutefois trop simple d'opposer les intuitions comme des options concurrentes où l'une dominerait sur l'autre. Il est en effet possible qu'elles répondent à des besoins et à des situations différentes. Dans la plupart des études que nous avons citées dans ce travail et qui impliquent des sujets en âge de parler, ces derniers connaissent ou sont informés des fins et des fonctions, ou des causes et des intentions, ou encore des propriétés et des usages des objets pour lesquels ils doivent choisir une explication, un nom ou une catégorie. En général, ils ne sont donc pas confrontés à des objets dont ils ne savent rien. Or, dans une telle situation, l'intuition étiologique serait prise en défaut. En effet, penser que la fonction d'un artefact ou d'un trait biologique dépend seulement de son histoire est à peu près inutile quand on ne connaît pas l'histoire en question. Et il est peu probable que les partisans de l'approche étiologique demeureraient impassibles s'ils se trouvaient face à face avec un pseudo-lion issu d'un accident cosmique en se disant qu'il n'y a pas de danger parce que ses griffes et ses dents (ou ce qui leur ressemble) n'ont pas la fonction de déchirer des proies car elles n'ont pas l'histoire qu'il faut. On peut au contraire parier que l'autre intuition, celle qui s'appuie sur les propriétés et les capacités actuelles des objets, leur commanderait de courir et de chercher un refuge.

Face à un animal qu'ils ne connaissent pas ou peu, les enfants d'âge préscolaire, en particulier vers l'âge de 5 ans, sont quant à eux capables d'inférer la fonction de leurs traits à partir de leur similitude avec ceux d'autres animaux, puis d'utiliser cette information pour en tirer des conclusions quant à leur comportement (Kelemen, Widdowson, Posner, Brown, & Casler, 2003). Par exemple, si on leur montre l'image d'un vison de Sibérie et d'un oiseau de mer (fou à pieds bleus) en leur disant que le premier passe beaucoup de temps sur terre et le second dans l'eau, et si on leur montre ensuite l'image d'une loutre en leur demandant si c'est un animal terrestre ou aquatique, les enfants sont capables de comprendre le lien fonctionnel entre les pattes palmées de l'oiseau et son comportement pour en inférer que la loutre est également un animal aquatique, ses pattes étant elles-aussi palmées, malgré la ressemblance générale entre la loutre et le vison et leur appartenance à une même catégorie (Mustélidés). L'inférence s'appuie donc sur une similitude structurelle entre les membres inférieurs des deux animaux et pas sur la connaissance de leur histoire[159].

159 On ne peut pas écarter la possibilité que l'intuition historique joue un rôle dans cette inférence. Les enfants pourraient en effet penser que les pattes palmées sont pour se déplacer dans l'eau, que c'est ce pour quoi elles sont faites, et que par conséquent les animaux avec des pattes de ce type sont faits

Dans certains cas, le fait de connaître la fonction historique d'un objet, ce pour quoi il est fait, s'avère même contre-productif. Nous avons vu précédemment qu'au cours du développement, les enfants tendent rapidement à assigner une fonction à chaque objet et évitent de l'employer pour un autre usage. Or, cela peut parfois constituer un obstacle pour trouver la solution d'un problème nouveau. Defeyter et Gelman (2003) ont en effet montré que les enfants de 6 et 7 ans ont plus de mal à résoudre un problème nouveau que ceux de 5 ans lorsque la solution implique d'utiliser un objet nouveau pour autre chose que sa fonction. Les plus jeunes sont capables de voir les propriétés fonctionnellement pertinentes de chaque objet et de raisonner en termes de relations moyens-fin indépendamment des fonctions dont on leur a fait la démonstration. Les plus âgés semblent quant à eux bloqués par leur connaissance : le fait de savoir qu'un objet a été conçu pour une fonction donnée les empêche de voir ses autres usages possibles. En revanche, quand ils n'en connaissent pas la fonction, leur performance (taux de réussite et vitesse d'execution) est meilleure que celle des jeunes. L'une des conclusions des auteurs est que la conception d'un artefact en termes de *design* (ce qui va ensuite devenir le *design stance* de l'adulte) commence à se mettre en place à partir de l'âge de 6 ans et limite les attributions fonctionnelles à une seule, celle pour laquelle il a été originellement conçu.

Ce qui pour nous est particulièrement intéressant dans cette étude, ce n'est pas le prétendu changement conceptuel avec l'âge, lequel viendrait encore une fois appuyer l'idée que le raisonnement historique finit par dominer d'autres modes de raisonnement plus primitifs, mais plutôt le fait que les enfants n'éprouvent pas de difficulté particulière à résoudre un problème *nouveau* quand ils ne peuvent pas recourir à l'histoire. Et nous soulignons le mot « nouveau » pour indiquer une différentiation des compétences selon la situation, car si la représentation historique n'est pas appropriée pour raisonner sur des problèmes nouveaux mais finit tout de même par s'imposer dans la pensée téléofonctionnelle de l'adulte, c'est vraisemblablement parce qu'elle est appropriée pour raisonner sur d'autres types de problèmes, à savoir ceux qui ne sont pas nouveaux. C'est une question d'économie cognitive. S'il fallait redécouvrir le feu et la roue à chaque fois qu'on en a besoin, l'humanité ne serait pas sortie des cavernes. En associant à chaque outil une fonction déterminée, c'est-à-dire un type de problème pour lequel cet outil est la solution, on peut résoudre des problèmes typiques très facilement, presque sans y penser, puisqu'il suffit de se souvenir de l'outil adéquat. Nous savons par exemple que le four à micro-ondes permet de chauffer très rapidement des aliments et des liquides, mais la plupart des gens ne connaissent pas le mécanisme causal qui en est responsable, pas plus qu'ils ne connaissent le

pour aller dans l'eau.

fonctionnement d'un moteur à explosion ou d'un ordinateur. Ce qui compte pour eux, c'est de connaître la fonction pour laquelle ils ont été conçus et la manière de les utiliser correctement.

Par contre, lorsque le problème est nouveau, on doit recourir au système de représentation téléologique, c'est-à-dire aux mécanismes cognitifs qui nous permettent de comprendre les relations moyens-fin en tenant compte des propriétés des objets et des contraintes de la situation. C'est une compétence que les jeunes enfants partagent avec le poulpe, le corbeau et le chimpanzé, mais aussi avec Archimède, Léonard de Vinci et Thomas Edison. Ce n'est donc pas un mode de raisonnement primitif, immature et infantile, que le raisonnement historique et le *design stance* viendraient dépasser à l'âge adulte.

Outre la recherche des solutions d'un problème donné, la pensée téléologique permet aussi d'adopter la démarche inverse qui consiste à partir d'un objet ou d'une propriété pour envisager les problème dont ce pourrait être la solution. C'est la base des explications fonctionnelles en biologie et dans d'autres domaines. Par exemple, à quoi peuvent bien servir les rayures des zèbres ? À se camoufler dans la végétation, déstabiliser le système visuel des prédateurs, réduire la différentiation des individus au sein du troupeau, réguler leur température ou encore limiter les piqûres d'insectes comme la mouche tsé-tsé ? Chacune de ces hypothèse a fait l'objet d'études et de publications scientifiques qui confirment par exemple que les rayures difficultent l'estimation de la vitesse et de la position des proies par un prédateur (Stevens, Yule, & Ruxton, 2008) et qu'elles sont peu attractives pour les taons parce qu'elles provoquent une modulation de la polarisation de la lumière (Egri et al., 2012).

Les explications précédentes répondent à la question posée en montrant comment les rayures apportent effectivement des solutions à plusieurs problèmes que rencontrent les zèbres. Elles sont compatibles avec l'analyse fonctionnelle de Cummins où la fonction d'un trait (les rayures) est une capacité de ce trait (l'altération de la polarisation de la lumière) qui contribue à expliquer une capacité plus générale de l'organisme (éviter les piqûres d'insectes, survivre dans la savane). En revanche, elles ne répondent pas directement à la question de la présence, à savoir : pourquoi les zèbres ont-ils des rayures ?

La pensée historique, quant à elle, dicte que la fonction des rayures est ce pour quoi elles ont été crées. Dans le cadre de la théorie darwinienne, on dira que leur fonction est ce pour quoi elles ont été sélectionnées. Le problème est que nous ne savons toujours pas si elles sont le produit de la sélection naturelle (Larison et al., 2015). Nous ne savons donc pas si elles ont une fonction au sens de la conception étiologique. Et quand bien même nous le saurions, nous pourrions ne pas connaître le facteur de sélection (thermorégulation ?), ni le mécanisme causal en vertu duquel ils remplissent leur fonction (formation de tourbillons d'air entre les franges sombres et claires ?), ni comment ce

mécanisme contribue effectivement à la *fitness* de l'organisme (évacuation de la chaleur superficielle ?), a fortiori si la notion de *fitness* est entendue comme succès reproductif historiquement avéré. En d'autres termes, nous pourrions expliquer pourquoi les zèbres ont des rayures sans savoir pour autant à quoi elles servent.

On comprend bien à travers cet exemple que les intuitions et les démarches qui sous-tendent les conceptions étiologique et dispositionnelle sont à la fois indépendantes et complémentaires. Elles coïncident dans la notion de « raison d'être » que nous avons empruntée à Larry Wright. Si les rayures des zèbres ont une raison d'être, cela veut dire que leur existence n'est pas fortuite, mais qu'il y a quelque chose qui la justifie au-delà de ses causes ; une justification liée à leurs effets. C'est-à-dire que leur existence est liée à leurs conséquences. Et peu importe qu'elles soient le produit d'un créateur intelligent ou de la sélection naturelle. Dans le premier cas, elles existent parce que le créateur désire leurs conséquences ; pour lui, elles sont une fin ou contribuent à une fin. Dans le second, elles existent parce qu'elles contribuent à la *fitness* des organismes porteurs et contribuent ainsi à leur propre existence et reproduction. Or, pour comprendre comment cette contribution est possible, il faut procéder à une analyse fonctionnelle à la Cummins. Par exemple, si la fonction des rayures est la thermorégulation, alors il faut expliquer comment elles sont effectivement capables de modifier la température corporelle et comment cette capacité contribue à la survie des zèbres dans la savane. L'explication fonctionnelle complète de ce trait requiert donc l'articulation des réponses à deux questions différentes, l'une portant sur son étiologie, l'autre sur ses capacités systémiques.

Ce que nous avons tâché de montrer aux CHAP. I & II et dans le reste de ce travail, c'est que les conceptions historiques et systémiques s'appuient sur des intuitions différentes qui capturent, chacune de son côté, des aspects importants du concept de fonction tel qu'il est employé en biologie et dans d'autres domaines. Nous comprenons maintenant que ces intuitions ont des sources psychologiques différentes qui les justifient et qui les limitent. Elles ne répondent pas aux mêmes questions ni ne correspondent aux mêmes besoins. La représentation téléologique est particulièrement adaptée à la compréhension des relations moyens-fin, c'est-à-dire à l'interprétation des actions téléologiques et des relations fonctionnelles, ainsi qu'à la résolution des problèmes nouveaux. L'intuition historique, pour sa part, conjointement avec la pédagogie naturelle, est particulièrement adaptée à la résolution de problèmes récurrents et à la transmission des connaissances culturelles — en particulier celles qui sont cognitivement opaques, c'est-à-dire inaccessibles à la pensée téléologique. Elles ne sont donc ni contradictoires ni incompatibles, mais largement complémentaires. Les tentatives d'unification des approches étiologique et systémique sont par conséquent justifiées du point de vue psychologique.

L'une de ces tentatives est celle de Philip Kitcher qui s'articule autour de la notion de *design* (CHAP. III). D'après cet auteur, les traits fonctionnels sont des réponses (ou contribuent à répondre) aux pressions sélectives de l'environnement ; ils ont été conçus pour cela. D'un côté, sa définition est liée à la pensée historique et à la conception étiologique dans la mesure où il considère que la fonction d'un trait est ce pour quoi il a été conçu et que la source de son *design* — mais pas forcément sa cause — est la sélection naturelle. De l'autre, elle est liée à la pensée téléologique et à la conception dispositionnelle dans la mesure où il considère que la fonction d'un trait est son rôle causal dans la réponse de l'organisme à une pression sélective et peut être déterminée à partir d'une analyse fonctionnelle.

Par exemple, pour attribuer une fonction aux rayures du zèbre, il faut identifier les « problèmes » environnementaux dont ce trait pourrait être ou faire partie de la « solution », comme la chaleur, les prédateurs, les insectes, et examiner si un ou plusieurs d'entre eux sont à l'origine de son apparition ou de sa conservation. Dans le cas contraire, on ne pourrait pas leur attribuer de fonction, quelle que soit leur utilité. De même, pour attribuer des fonctions aux rouages rouillés du mécanisme d'Anticythère et aux éléments de la prétendue « batterie babylonienne », il faut identifier les fins dont ces artefacts et leurs parties sont les moyens les plus efficients, puis examiner si ces fins ont effectivement motivé leur construction. Le problème est que leurs intentions nous étant inconnues, nous devons nous contenter d'une inférence vers la meilleure explication.

L'explication fonctionnelle de ces artefacts antiques associe l'archéologie à la rétro-ingénierie dont nous avons vu qu'elle s'appuie principalement sur le critère d'optimalité. La défense que nous faisions de ce critère et nos remarques quant à ses limitations (voir CHAP. III, SECT. 4 et CHAP. VI, SECT. 3) prennent tout leur sens à la lumière de la psychologie cognitive. Le principe d'action rationelle ou efficiente, au cœur de la pénsée téléologique, est en effet lui-même un principe d'optimalité, et nous venons de voir qu'il n'est pas suffisant pour attribuer une fonction à un artefact ou à un trait biologique, car il doit être complété par la pensée historique. La rétro-ingénierie est le reflet méthodologique de cette complémentarité psychologique.

Chapitre ix

Comment la téléologie peut-elle être scientifique ? (I)

À partir du XVII^e^ s., la physique est entrée dans le sûr chemin de la science, suivant l'expression de Kant. Elle est devenue pour cette raison un modèle de référence pour d'autres disciplines ainsi que pour la réflexion philosophique sur les sciences en général. Or, deux des principales notions employées depuis cette époque pour décrire et expliquer le monde physique sont celle de causalité, liée au mécanisme et entendue au sens de la cause efficiente d'Aristote, et celle de loi, qui établit une relation mathématique non nécessairement causale entre des grandeurs mesurables. Lorsque les philosophes se sont penchés sur la question de l'explication scientifique, au cours du XX^e^ siècle, plusieurs se sont donc naturellement tournés vers ces deux notions. Il s'agissait alors notamment de savoir si les sciences sont explicatives ou pas, de définir le concept d'explication scientifique et de comprendre comment les sciences spéciales pouvaient s'ajuster à un modèle général basé sur des lois ou sur des causes. C'est en partie pour répondre à ces questions qu'est né le débat sur les fonctions dont nous avons étudié les principales approches, car les explications de la biologie, comme celles de la psychologie et des sciences sociales, devaient s'adapter à un modèle inspiré de la physique. Cette idée, de nombreux chercheurs de ces disciplines l'ont eux-mêmes assumée :

> Pour la très grande majorité des enseignants et des chercheurs, le modèle par excellence de la science est celui élaboré par la physique. Le néo-positivisme, en toutes ses versions, demeure la philosophie implicite des chercheurs de terrain en sciences humaines. L'imprégnation majeure est assurée par le biais de la méthodologie, réduite le plus souvent aux simples procédés techniques, voire aux techniques d'enquête. Certes, ces mêmes chercheurs ne méconnaissent point le « verstehen », mais ils s'en méfient. Par rapport à l'explication, la compréhension ne produirait aucune véritable connaissance.

> Au mieux, elle confirmerait le déjà connu. Elle n'assurerait jamais la vérification empirique. Par conséquent, le type d'objectivité par excellence demeure celui de la physique. (Busino, 1988, p. 180-181)

Les sciences spéciales portent sur des objets très divers et elles les abordent depuis de multiples perspectives. Dans le seul domaine de la biologie, on compte plusieurs formes d'explication différentes ou, du moins, des pratiques explicatoires hétérogènes (voir par exemple Braillard & Malaterre, 2015). Les différentes conceptions du concept de fonction pourraient elles-mêmes correspondre à des usages différents parmi les sous-disciplines de la biologie. Vouloir toutes les juger à l'aune des explications de la science fondamentale, c'est défendre un monisme méthodologique et adopter une démarche réductionniste :

> [I]l y a un autre réductionnisme, plus subtil et plus radical […] qui repose non plus sur une sorte de *monisme ontologique*, à savoir sur l'idée que la réalité est d'un seul type, le type physique, et donc qu'est fondamentale la science qui traite de cette « base » de la réalité. La nouvelle forme de réductionnisme repose plutôt sur un *monisme méthodologique*. Cela signifie que l'on est prêt à reconnaître à toute science le droit de s'occuper de son objet spécifique, pourvu que la méthode qu'elle utilise soit la *vraie méthode* scientifique. (Agazzi, 1988, p. 14)

Dans ce contexte, nous commencerons par essayer de montrer que les explications physiques ne sont pas nécessairement incompatibles avec la téléologie. Ensuite, nous tâcherons de justifier la téléologie en réfutant certaines des objections que l'on peut lui opposer. Finalement, après avoir réfuté l'accusation d'anthropomorphisme, nous examinerons trois types de contributions que la téléologie peut apporter à la biologie d'un point de vue scientifique : des prédictions, des catégories et des explications spécifiques.

1. Les explications téléologiques ne mentionnent pas les causes

Prenons l'exemple séminal de Gergely et al. (1995). Nous voyons un petit cercle s'avancer vers un rectangle, le contourner, puis continuer en direction d'un grand cercle au contact duquel il s'arrête. Nous expliquons cela en disant que le petit cercle contourne le rectangle *pour* s'approcher du grand cercle. Ce faisant, nous ne prétendons pas que le contact ou la proximité entre les deux (l'état final) soit la cause du comportement observé. Et nous n'attribuons pas non plus — ou pas nécessairement — d'états mentaux intentionnels aux figures géométriques qui composent la scène. Si nous le faisions, nous pourrions l'interpréter à la manière de Davidson, en donnant du comportement une explication psychologique

et en considérant celle-ci comme une forme d'explication causale. Mais puisque l'attitude téléologique est indépendante de la théorie de l'esprit et vraisemblablement antérieure, la finalité qu'on attribue au comportement n'est donc ni une cause finale, ni une représentation mentale. Curieusement, nous sommes tentés de nous demander ce que c'est, en termes ontologiques, comme si la causalité et l'intentionnalité étaient transparentes de ce point de vue, alors qu'à l'évidence elles le ne sont pas. En termes psychologiques, nous nous demanderons plutôt ce que signifient les explications téléologiques, ce que nous voulons dire par là et ce qui les distingue des explications causales.

Prenons un exemple plus simple. Nous voyons un objet *A* en mouvement. Nous pourrions nous attendre à ce qu'il explose, qu'il change de forme, qu'il se divise en cinq, qu'il disparaisse soudainement ou quoi que ce soit d'autre, mais nous optons normalement pour une solution plus simple qui est qu'il continue son mouvement sans autres changements, de façon rectiligne et uniforme. Il n'y a rien de particulier à expliquer, si ce n'est éventuellement sa cause (en supposant qu'un mouvement ait toujours besoin d'une cause, ce qui n'est pas le cas en physique).

Si des situations similaires se répétaient, on pourrait réaliser tous types de prédictions, mais les deux les plus simples sont intuitivement les suivantes. La première s'appuie sur l'idée que le mouvement de *A* n'a pas de cause ni de raison particulière, de sorte que nous n'avons pas à nous attendre à ce qu'il adopte une direction plutôt qu'une autre, par conséquent :

(P_1) *A* se déplacera dans n'importe quelle direction.

La seconde s'appuie sur l'idée que le mouvement de A a une cause (que nous ignorons) et que les mêmes causes produisent les mêmes conséquences :

(P_2) *A* se déplacera dans la même direction.

Introduisons un deuxième objet *B*. Si l'une ou l'autre des deux prédictions précédentes est correcte, alors il n'y aucune raison de penser que *A* entrera en contact avec *B*. En effet, pour que cela se produise, il faudrait que *B* se trouve par hasard sur la trajectoire de *A*, laquelle sera soit la même soit une autre quelconque. Or, le nombre de directions possibles est virtuellement infini. Par conséquent, la rencontre des deux objets est extrêmement improbable. Imaginons cependant qu'elle ait lieu. On peut continuer à penser qu'il s'agit d'un hasard, aussi improbable soit-il, ou lui chercher une autre explication raisonnable. Nous nous limiterons ici aux explications causales et téléologiques pour essayer de comprendre ce qu'elles signifient, ce qu'elles impliquent et en quoi elles se ressemblent ou se distinguent.

Une explication causale simple du phénomène observé pourrait être la suivante :

> (E_1) *A* se déplace vers *B* ***parce que*** *A* subit l'attraction magnétique de *B*.

Une explication téléologique du même phénomène pourrait être :

> (E_2) *A* se déplace vers *B* ***pour*** entrer en contact avec lui.

Toutes deux établissent une relation entre les deux objets, mais tandis que la première s'aventure à formuler une hypothèse quant à la nature de la relation, la seconde reste vague sur ce point. La première fait référence à la situation initiale, la seconde à la situation finale. L'une et l'autre permettent d'inférer le résultat du déplacement de *A*, à savoir son contact avec *B*, avant qu'il ne se produise. Il convient de noter qu'ici rien ne nous porte à croire que l'explication téléologique implique une cause finale ou rétrograde[160], ni une cause surnaturelle, ni une absence de cause, ni une attribution intentionnelle, ni une prédétermination de l'avenir, ni encore une orientation ou une direction quelconque de l'ordre des choses. En effet, rien ne nous empêche de compléter l'explication téléologique par une explication causale :

> (E_3) *A* se déplace vers *B* ***pour*** entrer en contact avec lui ***parce que*** *A* subit l'attraction magnétique de *B*.

Ceci dit, il nous semble qu'une telle formulation rend superflue la partie téléologique, car nous ne lisons dans E_3 rien de plus ni rien de différent par rapport à E_1. Peut-être parce que les explications téléologiques se caractérisent justement par le fait de ne pas mentionner les causes. Suivant cette interprétation, on aboutirait à l'explication suivante qui n'est qu'une reformulation plus explicite de E_2 :

> (E_4) *A* se déplace vers *B* ***pour*** entrer en contact avec lui (quelles qu'en soient les causes).

160 Le concept de cause ou de relation causale semble intimement lié à l'idée selon laquelle le temps va dans une direction déterminée et irréversible, du passé vers le futur. Cette association entre la causalité et la flèche du temps est au cœur des accusations de causalité rétrograde dirigées contre la téléologie, alléguant qu'elle implique une influence contre nature de l'avenir sur le passé. Accusation d'autant plus curieuse qu'elle a été forgée par certains des fondateurs de la mécanique classique. Curieuse parce que l'exemple prototypique d'une relation causale est le choc élastique entre deux boules de billard, alors même que les équations qui le décrivent sont réversibles dans le temps. Il n'y a donc pas, du point de vue de la physique classique, de direction temporelle privilégiée.

Cela étant, on peut légitimement se demander dans quelle mesure E_4 constitue une explication authentique et en quoi elle se distingue d'une description comme :

> (*Description*) *A* se déplace vers *B* jusqu'à entrer en contact avec lui (quelles qu'en soient les causes).

2. Les explications téléologiques expliquent-elles quelque chose ?

En premier lieu, on peut dire simplement que E_4 implique une relation (a priori indéterminée) entre *A* et *B*, mais c'est beaucoup trop vague. On peut alors ajouter la précision suivante : E_4 implique que le déplacement de *A* en direction de *B* n'est pas accidentel mais se trouve conditionné d'une façon ou d'une autre par son résultat. Ou encore : E_4 implique que la relation entre le déplacement et son résultat est de type moyens–fin. Aucune de ces implications n'est présente dans la description. Or, elles permettent de réaliser certaines inférences et vérifications que la description seule ne permet pas.

Par exemple, on peut penser que si *B* n'était pas présent ou venait à disparaître, alors *A* ne se déplacerait pas ou cesserait de le faire. De plus, si *A* déviait de sa trajectoire à l'approche de *B*, de sorte qu'ils finissent par ne pas entrer en contact, ou si *B* changeait de position et *A* conservait la sienne, avec le même résultat négatif, l'explication serait faussée. En outre, si l'on plaçait un obstacle entre les deux objets, on pourrait s'attendre à ce que *A* modifie sa trajectoire de telle manière qu'il puisse tout de même atteindre *B*. En effet, les relations moyens–fins impliquent une certaine variabilité des moyens pour atteindre la même fin. Par conséquent, l'explication E_4 inscrit un cas particulier dans un cas plus général, ce que la simple description ne fait pas.

Par ailleurs, l'explication téléologique impose certaines restrictions à l'équifinalité et à la multiréalisabilité, car tous les moyens ne se valent pas. Si *A* réalisait des cabrioles ou s'éloignait puis se rapprochait à nouveau, E_4 ne rendrait pas compte de ces comportements. Dans le cas qui nous occupe, parmi toutes les trajectoires possibles de *A* vers *B*, il y en a une qui se distingue de toutes les autres, car elle implique la moindre distance et le moindre temps ; c'est la ligne droite. Si la théorie de Gergely et Csibra est correcte, alors les explications téléologiques ne rendent compte que des comportements qui vérifient le principe d'action efficiente, c'est-à-dire ceux qui, dans un contexte situationnel donné, constituent le moyen le plus efficient disponible pour atteindre une fin. Par conséquent, E_4 ne constitue une explication valable du déplacement de *A* que si ce dernier est le plus efficient pour atteindre *B*. Mais étant donné que le

déplacement le plus efficient dépend du contexte et que celui-ci n'est pas stipulé, ce que fait E_4 c'est ramener le cas particulier du déplacement en ligne droite de *A* vers *B* à une « loi » générale disant que *A* se déplacera toujours vers *B* de la manière la plus efficiente.

Depuis cette perspective, on peut penser que l'explication téléologique correspond au modèle d'explication déductif nomologique (D-N) où une loi de couverture subsume les cas particuliers et permet de les prédire à l'aide de conditions initiales appropriées. En effet, à partir d'une situation initiale quelconque, on peut déterminer les moyens les plus efficients pour arriver à une situation finale donnée ; quels que soient les obstacles que l'on place entre *A* et *B*, on peut calculer le comportement le plus efficient du premier pour parvenir au second.

Cette approche de la téléologie a été défendue notamment par Nancy Cartwright (1986), pour qui les explications téléologiques en biologie apportent le même type de compréhension que les lois de couverture en physique et ont la même légitimité scientifique. Cette dernière dépend de la manière dont la théorie à laquelle appartiennent ces lois parvient à sauver les phénomènes :

> « *p is the purpose of behavior b* is empirically warranted when it is an essential claim of a total explanatory theory that best saves the phenomena. […] Here is kind of teleology that is no efficient causality in disguise. Its explanatory power is coextensive with its empirical warrantability, and how great that is depends on what kind of theoretical organization works best in practical biology. » (Cartwright, 1986, p. 209)

Les propositions E_2 et E_4 feraient donc partie d'une théorie plus générale dont la validité dépend de son adéquation empirique, c'est-à-dire de la manière dont elle rend compte des observations et du succès des prédictions qu'elle permet de formuler. Supposons par exemple que *A* soit un rat et *B* une source de nourriture. Il ne fait aucun doute que l'on peut expliquer et prédire le comportement du rat en présence de nourriture dans de nombreuses situations différentes : si rien ne les sépare, il ira en ligne droite ; s'il y a un obstacle, il le contournera… Dans chaque cas particulier, l'explication et la prédiction dérivent d'une théorie plus générale du comportement de cet animal. L'explication est alors valable non seulement pour cet individu particulier, mais pour tous ceux de son espèce. La théorie en question est sans doute dérivable à son tour de théories et de principes plus généraux portant sur le comportement de beaucoup d'autres animaux. On devrait ainsi pouvoir expliquer, à partir des mêmes principes, le comportement d'une colonie de fourmis en présence d'un morceau de pain ou d'un prédateur en présence d'une proie. On pourrait peut-être même expliquer, à partir de principes encore plus fondamentaux, le mouvement de bactéries nageant dans un gradient

de sucres (chimiotactisme) ou l'orientation d'une plante verte vers une source de lumière.

Quand on attribue aux phénomènes biologiques des fins et des fonctions, comment savoir si ces attributions sont correctes ? D'après cette auteure, la réponse est la même que pour les énoncés théoriques de la physique. Les conditions de validité des théories biologiques et physiques seraient en effet les mêmes, et la manière dont elle les formule n'est pas sans rappeler la conception de l'explication de Philip Kitcher en termes d'unification théorique :

> « When will such functional ascriptions be correct? As with any theoretical claim, when enough of its consequences are borne out. The theory is better the wider the variety of consequences, the greater their precision, the more roads for the further investigation opened, the more problems solved, and so forth. » (Cartwright, 1986, p. 208)

Cartwright insiste par ailleurs sur le fait que les explications téléologiques ne sont pas causales, pas plus que les explications théoriques de la physique. Il faut en effet distinguer entre l'explication causale d'un phénomène, comme celle du mouvement brownien à partir de l'agitation des atomes et des molécules d'un fluide, et les explications théoriques qui proposent des modèles et des lois mathématiques de ce même phénomène. En biologie, l'évolution par sélection naturelle permet de donner une explication causale à de nombreux phénomènes biologiques, y compris les traits et les comportements fonctionnels, mais alors la portée de l'explication et ce qui fait qu'elle soit valable s'en trouvent limités : pour qu'une affirmation comme « la fonction de X est Z » soit empiriquement correcte, il faut que Z soit une conséquence de X et qu'elle ait été historiquement sélectionnée. Si les fonctions ne sont pas des explications causales déguisées, alors elles ne dépendent pas de l'évolution par sélection naturelle et leur portée comme leurs conditions de validié s'en trouvent élargies.

Nous reviendrons plus longuement sur la justification de la valeur explicative des explications téléologiques en biologie dans le CHAP. XI, SECT. 5 à 7.

3. Comparaison des deux types d'explication

Comparons maintenant les explications causales et téléologiques depuis une autre perspective. L'une des différences les plus notables que l'on remarque est le rapport qu'elles établissent entre le particulier et le général. En ce qui concerne E_1, nous avons émis l'hypothèse d'une attraction magnétique pour expliquer le comportement de A, mais la nature

exacte de la cause importe peu pour la comparaison que nous voulons établir. Ce que nous voulons souligner, c'est que E_1 n'établit pas une relation entre deux objets particuliers mais entre certaines propriétés de ces objets. Nous supposons par exemple que B est un dipôle magnétique (un aimant) et que A est un objet paramagnétique (une bille en fer), ou vice-versa ; mais peu importe leur identité, puisqu'on pourrait les remplacer par deux autres objets quelconques ayant les mêmes propriétés pertinentes. De manière très générale, E_1 revient à dire : (1) tous les objets ayant certaines propriétés causales se comportent nécessairement d'une certaine façon dans certaines circonstances ; (2) *A* et *B* ont les propriétés en question et se trouvent dans ces mêmes circonstances ; par conséquent, (3) *A* et *B* doivent nécessairement se comporter de cette façon, laquelle correspond au déplacement que l'on observe.

L'explication téléologique E_4, quant à elle, dans le cas présent, ne mentionne aucune cause ni aucune propriété des objets qui permettrait de les remplacer par d'autres ayant les mêmes propriétés pertinentes de manière à conserver la relation explicative. La seule chose que nous sachions ou que nous supposions à propos de *A* et de *B* est relatif à leur comportement, et comme nous ignorons ce qui le cause ou le motive, nous ignorons aussi s'il leur est spécifique ou partagé par d'autres. Autrement dit, nous ne savons pas si en remplaçant *A* par *A′* ou *B* par *B′* nous observerions le même déplacement. E_4 permet de généraliser le comportement des objets dans différentes situations, mais contrairement à E_1 elle ne permet pas de faire abstraction des objets eux-mêmes. L'explication causale fait ici référence à des objets quelconques ayant des propriétés particulières. L'explication téléologique, quant à elle, fait ici référence à des objets particuliers ayant des propriétés quelconques[161].

Une autre différence notable entre les deux explications porte sur leurs prédictions respectives. Toutes deux conduisent à prédire que si *B* change de position ou se déplace, *A* se déplacera aussi de façon à arriver au même résultat, mais les deux n'ont pas la même précision. Si le déplacement de *A* est causé par l'attraction magnétique de *B*, alors il est soumis à une accélération constante que l'on peut calculer et que l'on doit pouvoir observer. L'explication causale donne donc des détails précis qui s'appuient sur des lois physiques et qui permettent de la réfuter ou de la corriger. L'explication téléologique est plus vague, du moins en apparence. Nous reviendrons sur ce point.

161 Cela ne veut pas dire que les explications téléologiques ne soient applicables qu'à des objets particuliers. Ici, nous ne savons pas si le comportement observé est propre à ces objets-là ou s'il est partagé par d'autres. Il pourrait par exemple être partagé par tous les membres d'une même catégorie. Si *A* est un rat et *B* une source de nourriture, on comprend rapidement que la même explication est extensible à n'importe quel rat et à diverses sources de nourriture.

On constate aussi qu'elles conduisent à des prédictions différentes dès que la situation se complique un peu. Ajoutons par exemple un troisième objet *C* faisant obstacle entre les deux précédents. Le raisonnement téléologique nous dit que *A* va contourner *C* pour atteindre *B*. Le raisonnement physique nous dit plutôt que *A* se déplacera vers *B* en ligne droite, butera sur l'obstacle et rebondira ou restera sur place. Dans l'expérience de Gergely et ses collègues, le comportement de *A* vérifie la prédiction téléologique.

Ce dernier comportement peut aussi recevoir une explication causale. En conservant l'idée d'une attraction magnétique entre *A* et *B*, on pourrait supposer par exemple une répulsion magnétique de moindre intensité entre *A* et *C* (diamagnétisme) ou une zone d'exclusion totale du flux magnétique comme celle que créent les matériaux supraconducteurs (effet Meissner).

De nouveau, on peut tester les prédictions auxquelles conduisent respectivement les explications téléologique et causale en variant les conditions expérimentales. Par exemple, déplaçons un peu l'obstacle *C* de telle sorte que *A* puisse, en le frôlant, aller vers *B* en ligne droite[162]. Le raisonnement téléologique nous dit que c'est ce qui va effectivement se produire. Le raisonnement physique nous dit que la trajectoire de *A*, au voisinage de *C*, sera altérée d'une certaine façon par la perturbation du champ que crée le diamagnétisme parfait de celui-ci.

Si ce dernier comportement était observé, alors l'explication téléologique précédente ne serait pas suffisante pour en rendre compte et il faudrait la compliquer, de même que nous avons compliqué précédemment l'explication causale. Dans tous les cas, les deux explications sont possibles, mais en fonction de la situation l'une s'avère plus simple ou plus plausible que l'autre. En outre, dans la mesure où elles conduisent à des prédictions différentes, on peut les tester indépendamment de leur complexité ou complication relative.

Maintenant, remplaçons *A* par un rat et *B* par une source de nourriture. Dans un espace dégagé, le rat ira vers la nourriture en ligne droite. Si nous changeons la nourriture de place, il changera de direction. Si nous plaçons un obstacle entre les deux, il contournera l'obstacle. Si nous plaçons un labyrinthe entre les deux, il résoudra le labyrinthe. Comparons maintenant les deux explications et leurs prédictions respectives. Du point de vue téléologique, peu importe que *A* soit une bille métallique ou un rongeur, et peu importe la complexité de la situation, car l'explication reste la même aussi longtemps que l'objet continue d'obéir au principe d'action efficiente : *A* se déplace vers *B pour* entrer en contact avec lui (et le manger). Par contre, d'un point de vue causal, la nature de l'objet est essentielle et la complexité de l'explication dépend directement de la

162 Cette variation expérimentale m'a été suggérée par Manouk Abkarian.

situation. Expliquer en termes strictement physiques les différents comportements du rat dans les différentes situations indiquées est alors sinon impossible, du moins impraticable.

Ce que nous voulons souligner ici, c'est que les explications causales ont l'avantage indéniable de la précision, mais que cela les rend plus lourdes et moins flexibles, surtout quand il s'agit de rendre compte du comportement complexe d'objets complexes dans des environnements complexes. Les explications téléologiques, pour leur part, sont imprécises parce qu'elles font abstraction des causes et des lois physiques qui régissent concrètement un comportement donné dans une situation donnée, mais c'est justement cette imprécision qui leur permet de rendre compte des différents comportements possibles d'un même objet dans des situations différentes en vertu de ce qu'ils ont en commun : le résultat attendu, le but visé.

Ce résultat attendu ou but visé n'est pas une cause finale, ni un objectif fixé par un agent surnaturel, ni une représentation mentale de l'objet lui-même. C'est le résultat que *nous*, en tant qu'observateurs, pouvons attendre du comportement de l'objet en vertu du but vers lequel *celui-ci* se dirige. Et cela vaut pour un organisme vivant, une bille métallique, un thermostat ou une étoile. Quelle que soit la nature de l'objet, on peut opter pour une interprétation téléologique de son comportement ou pour une interprétation physique, car ce sont deux des mécanismes cognitifs dont *nous* sommes naturellement dotés pour interpréter le monde qui nous entoure.

4. Principes d'optimalité

Pourquoi alors avons-nous tendance à expliquer les objets biologiques en termes téléologiques et les autres objets naturels en termes physiques ? Revenons à l'expérience du petit cercle, du grand cercle et du rectangle. D'après les chercheurs, les enfants adoptent l'attitude téléologique et voient le petit cercle comme un agent rationnel et pas comme un objet physique. En réalité, les deux interprétations sont en principe possibles. Nous avons vu que si le petit cercle représentait une bille métallique et le grand cercle un aimant, alors la bille se déplacerait en ligne droite en l'absence d'obstacle. Le fait que les enfants s'y attendent n'implique donc pas nécessairement une interprétation téléologique. En ce qui concerne l'autre comportement qui consiste à contourner l'obstacle, nous avons vu qu'il peut aussi sans aucun doute recevoir une explication physique, mais cette explication est loin d'être évidente. En effet, ce n'est pas le type de comportement que l'on observe habituellement dans la nature, hormis chez les êtres vivants, et ce n'est pas ce que l'on attend intuitivement d'un objet inerte. À vrai dire, les phénomènes magnétiques ne sont pas non plus très courants dans la nature, aussi

simples soient-ils. Par conséquent, si nous disposons dès la naissance de systèmes cognitifs différents et si à chacun de ces systèmes correspond un domaine d'objets naturels que l'on peut grossièrement diviser en physiques et biologiques, c'est parce que ces objets ont des comportements bien particuliers qui requièrent des stratégies interprétatives spécifiques.

C'est à partir de son comportement que les enfants classent un objet comme le petit cercle dans la catégorie des agents ou des objets inertes. Or, nous avons vu que les indices d'agentivité incluent le mouvement autonome (auto-initié, auto-propulsé et « libre »), l'équifinalité des variations (différentes actions, même résultat) et la capacité d'interagir de façon dynamique et contingente avec l'environnement. Cependant, nous avons vu aussi que pour Csibra, Gergely et leurs collègues (1999), l'attitude téléologique n'est pas un système interprétatif basé sur des indices, mais sur un principe. Si le petit cercle est considéré comme un agent, c'est parce que son comportement s'ajuste à ce principe ; parce qu'il réalise l'action la plus efficiente pour atteindre son but présumé (arriver près du grand cercle) étant donné les contraintes de la situation (le rectangle faisant obstacle). Et si l'on peut lui attribuer ce but, c'est parce qu'il n'y en a aucun autre qui vérifie le principe, c'est-à-dire aucun autre relativement auquel l'action observée soit la plus efficiente.

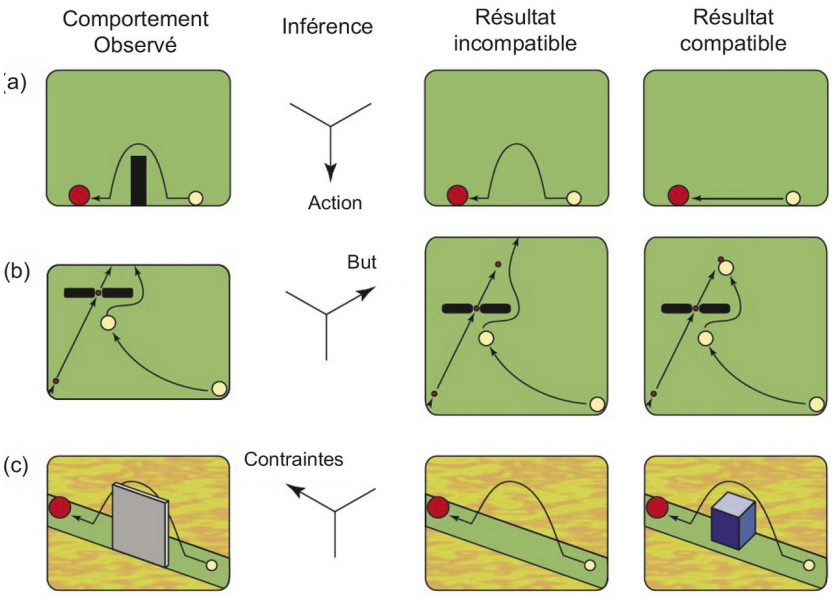

Figure 25: Principe d'action rationnelle. Le comportement d'un agent est dit rationnel lorsqu'il existe une relation efficiente entre l'action, le but et les contraintes. À partir de deux de ces éléments, il est alors possible d'inférer le troisième. (Source Gergely & Csibra, 2003).

Il s'agit donc d'un principe d'optimalité qui met en relation les trois éléments qui composent l'action et leur donne une signification téléologique. Le contexte physique, le comportement et l'état final deviennent respectivement les contraintes situationnelles, les moyens et la fin (Fig. 25). La relation établie entre ces éléments permet d'inférer intuitivement n'importe lequel d'entre eux à partir des deux autres. Des études ont en effet montré que lorsque l'un des trois est caché à la vue, les attentes et inférences des enfants de 12 mois sont conformes à l'application du principe (Csibra, 2003; Csibra, Bíró, Koós, & Gergely, 2003).

En physique, il existe aussi des principes d'optimalité qui mettent en relation les trois éléments de l'action et qui permettent d'inférer les uns à partir des autres, et nous allons voir que ces principes sont analogues au principe d'action efficiente.

Considérons la situation suivante, tirée des *Lectures on Physics* de Richard Feynman (Fig. 26). Sur la plage (milieu 1), un secouriste (A) voit un baigneur (B) en train de se noyer en mer (milieu 2). Quel trajet doit-il adopter pour le sauver ? Au premier abord, on pourrait penser qu'il doit aller en ligne droite, car c'est le chemin le plus court. Ce faisant, il parcourra la moitié du trajet sur la plage et l'autre moitié dans l'eau. Or, les déplacements dans l'eau sont nettement plus lents que sur la plage. Donc, en réduisant la distance à nager dans l'eau tout en allongeant celle à courir sur la plage, on peut racourcir le temps total pour arriver au but. Le trajet de plus efficient, dans ce cas, n'est pas celui qui minimise la distance, mais le temps de parcours. Et si quelqu'un observant le secouriste se demandait pourquoi il se comporte de cette façon, on pourrait lui donner une explication téléologique : c'est pour arriver plus vite.

Remplaçons maintenant le secouriste par un rayon lumineux, la plage par un gaz (air) et la mer par un milieu dense (eau, verre). Imaginons que le rayon lumineux aille de *A* vers *B*. Quel trajet va-t-il adopter ? La réponse à cette question a été apportée par Pierre de Fermat en 1657 en disant que la lumière se propage entre deux points en suivant la trajectoire qui minimise le temps de parcours. Cela implique que la forme du chemin optique d'un rayon lumineux est comparable à celle du trajet de

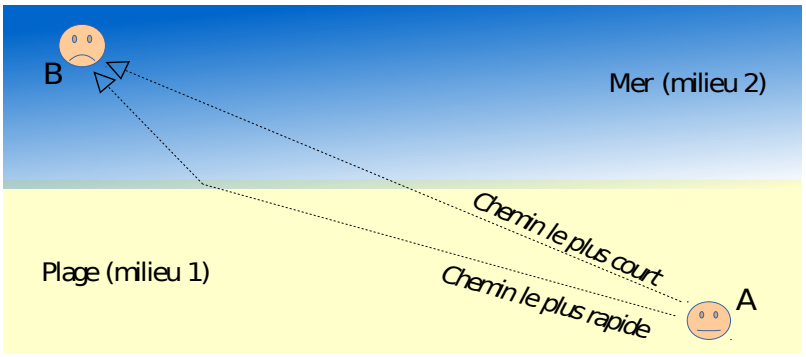

Figure 26: Principe de moindre temps. Exemple de Richard Feynman pour illustrer le principe : pour sauver un baigneur (B) en danger, le secouriste (A) doit adopter le trajet le plus rapide, pas le plus court, c'est-à-dire celui qui réduit le temps de parcours en fonction de la nature des milieux parcourus.

notre secouriste[163]. C'est le principe de moindre temps ou d'économie naturelle.

Historiquement, Fermat a tiré son principe d'optimalité à partir de considérations métaphysiques et théologiques, estimant — comme plusieurs de ses contemporains — que la nature agit toujours par les voies les plus simples et les plus courtes. Jean le Rond D'Alembert le critique durement dans son article « Causes finales » de l'*Encyclopédie*, arguant que l'usage des causes finales est stérile et dangereux, bien qu'il n'hésite pas à s'en servir lui-même quelquefois pour appuyer un principe de mécanique (Ru, 1994, p. 104). De plus, le principe de Fermat est téléologique aussi au sens où, pour déterminer le chemin optique le plus court d'un rayon de lumière, il faut définir son point d'arrivée avant qu'il n'y soit parvenu. Malgré tout, il permet de retrouver la plupart des résultats de l'optique géométrique et est considéré comme un jalon important de l'histoire des sciences modernes.

Près d'un siècle après Fermat et s'appuyant sur des considérations métaphysiques similaires, Pierre Louis Moreau de Maupertuis formula en 1744 le principe de moindre action selon lequel « *la Nature dans la production de ses effets agit toujours par les moyens les plus simples [et] lorsqu'il arrive quelque changement dans la Nature, la quantité d'Action employée pour ce changement est toujours la plus petite qu'il soit possible* ». Comme celui de Fermat, il est doublement téléologique, d'abord parce que d'après Maupertuis son principe « *laisse le Monde dans le besoin continuel de la puissance du Créateur et est une suite nécessaire de l'emploi le plus simple de cette puissance* », ensuite parce qu'il requiert deux points fixes pour calculer le trajet d'un mobile, un point de départ et un point d'arrivée.

Privés de leur soutien métaphysique, les principes de Fermat et de Maupertuis sont des conjectures indémontrables, *si ce n'est à partir d'autres principes*, car il n'y a aucune cause ni aucune raison connue en vertu de laquelle, dans un cas de réfraction comme celui de notre exemple, la lumière adopte le chemin de moindre temps et non celui de moindre distance ou n'importe quel autre chemin possible. Pour rendre compte du comportement de la lumière entre *A* et *B*, nous ne pouvons donc pas en donner une explication causale. Nous devons nous contenter de dire que la lumière adopte ce chemin « *parce que* c'est le plus rapide » ou « *pour* minimiser le temps de parcours », ce qui est équivalent.

Reformulé mathématiquement par Lagrange et Euler, le principe de moindre action allait permettre de redéfinir toute la physique de Newton. Une grande partie de la physique postérieure, dont l'électromagnétisme, la relativité générale et la mécanique quantique, peut aussi se déduire du

163 Cet exemple ainsi que d'autres éléments de l'argument qui suit à propos des principes d'optimalité en physique m'ont été inspirés par une conférence de Madjid Mesli (2013) sur le principe de moindre action.

même principe. De manière générale, les principes variationnels, qui reposent sur l'optimisation d'une grandeur, sont des outils fondamentaux et très puissants employés dans de nombreuses branches de la physique.

Cependant, il est important de souligner ici, pour éviter toute confusion, que les phénomènes physiques que l'on peut expliquer à l'aide de principes variationnels peuvent également recevoir une explication par des lois *locales*, lesquelles ne sont pas susceptibles d'une interprétation téléologique. En effet, on peut considérer que le comportement d'un système est déterminé à chaque instant par ses conditions locales ou causales, de sorte qu'il n'est pas nécessaire de faire appel à ses conditions finales. Ainsi, les lois de Descartes prévoient le chemin suivi par un rayon lumineux initial donné, tandis que Fermat se demande quel est le chemin effectivement suivi par la lumière pour aller d'un point à un autre. Les lois des Descartes sont locales, le principe de Fermat est global, mais les deux arrivent aux mêmes résultats. Il s'agit de deux formulations différentes d'une même théorie qui sont équivalentes dans la mesure où l'on peut dériver l'une à partir de l'autre, mais elles présentent chacune divers avantages et inconvénients.

En disant que ces explications physiques sont téléologiques au sens où elles impliquent un état final ou optimal du système, nous ne faisons pas une lecture ontologique ou réaliste de ces principes. La téléologie et la causalité ne sont pas entendues ici comme des modes de relation dans le monde, mais comme des modes de raisonnement sur le monde. Nous avons insisté au CHAP. VIII sur le fait que, du point de vue de la psychologie cognitive, elles ne sont pas concurrentes, mais complémentaires. André Lichnerowicz exprimait à peu près la même idée, mais du point de vue mathématique :

> « les mathematiciens ont étudié les systèmes differentiels (d'apparence causaliste) qui peuvent être considérés comme émanant d'un principe variationnel (donc à vocation finalisante); ils ont montré qu'il s'agissait d'une large classe de systèmes satisfaisant une propriété générale abstraite dite cohomologique et que, pour des raisons non moins générales, les systèmes aptes à la représentation du réel physique devaient posséder cette propriété. Le caractère universel des principes variationnels apparaissait en pleine lumière. Il est sans doute permis de dire que, selon le regard posé sur elle, toute théorie physique peut être considérée comme causaliste ou finaliste, alors qu'il s'agit d'une seule et même théorie. À ce niveau d'intelligence du réel, causalité et finalité apparaissent comme des notions inadéquates, ne pouvant servir, ici ou là, que de béquilles heuristiques. » (Lichnerowicz, 1987)[164]

164 Cité par Jean Mawhin (1998).

Les explications fondées sur des principes variationnels sont plus générales, plus flexibles, plus intuitives et plus simples. D'après Max Born, s'il devait y avoir une équation capable de décrire l'univers dans son ensemble, elle prendrait cette forme :

> Nous sommes encore loin de connaître la formule universelle de Laplace mais nous pouvons être convaincu qu'elle aura la forme d'un principe extrémal, non pas parce que la nature a une volonté, un but ou une économie, mais parce que le mécanisme de notre pensée n'a pas d'autre voie pour condenser une structure de lois compliquée en une brève expression. » (Born, 1939)[165]

Cette clarification étant faite, revenons à notre argument. Un ballon lancé depuis la plage vers la mer ira toujours en ligne droite de *A* vers *B*, suivant le chemin le plus court mais pas le plus rapide. Pourquoi le rayon ne fait-il pas de même ? Pourquoi, parmi tous ces chemins possibles, le rayon choisit-il le plus rapide et comment peut-il savoir à l'avance lequel d'entre tous serait le plus rapide ? Ces questions sont absurdes en ce qu'elles attribuent à la lumière des capacités de représentation et de décision qu'elle ne possède pas, mais elles illustrent bien les problèmes que pose l'interprétation de la téléologie. Nous sommes en effet portés à faire une lecture mentaliste ou finaliste de la préposition « pour » dans les explications téléologiques. Lorsque la chose expliquée est un rayon de lumière ou un système physique, nous évitons cette interprétation parce que *aujourd'hui* elle nous semble absurde (alors qu'elle ne l'était pas pour certains illustres scientifiques du XVII^e^ s.). Nous avons donc fini par laisser de côté les considérations métaphysiques des principes variationnels en physique, mais nous avons encore du mal à le faire en biologie.

Notre argument est simple : si nous pouvons laisser de côté les considérations métaphysiques pour la lumière et nous limiter à expliquer son comportement en montrant qu'il s'ajuste au principe de moindre temps, alors nous pouvons (nous devons) aussi le faire pour d'autres objets dont le comportement est lié à d'autres principes. Ainsi, si nous remplaçons le secouriste par un autre agent quelconque, représenté par exemple par une figure géométrique comme dans l'expérience de Gergely et Csibra, nous pouvons de même expliquer son comportement en montrant qu'il s'ajuste au principe d'action rationnelle ou efficiente. Dans un cas comme dans l'autre, nous pouvons dire que l'objet (rayon ou agent) suit ce trajet « *parce que* c'est le plus rapide » ou « *pour* minimiser le temps de parcours » sans lui attribuer pour autant une rationalité ni faire référence à ses états mentaux, ni à la Providence divine, ni à une cause finale.

165 Idem.

Les principes d'économie naturelle et d'action efficiente établissent des relations constantes entre les éléments de la situation et permettent par conséquent d'inférer les uns à partir des autres. En physique, il s'agit par exemple des conditions initiales, du comportement du système et de son état final. En téléologie, comme nous le disions plus haut, ce sont respectivement les contraintes, les moyens et la fin. Ce qui est rationnel, ce n'est pas l'objet, c'est-à-dire le système physique ou l'agent, mais la relation attendue entre ces trois éléments.

L'application du principe d'action rationnelle ou efficiente a par exemple donné lieu à une théorie (*optimal foraging theory*) qui vise à expliquer le comportement de recherche de nourriture d'un individu en termes de stratégie optimale (voir Pyke, 1984). Elle permet d'expliquer pourquoi les corbeaux de la côte nord-ouest du Canada choisissent soigneusement les bulots les plus grands et les plus lourds puis s'envolent pour les lâcher depuis une hauteur de près de 5 mètres sur les roches. Un chercheur a en effet calculé que c'est la hauteur optimale pour briser la coquille des mollusques en minimisant l'énergie nécessaire pour y arriver et en tenant compte des contraintes physiques (Zach, 1978). Cela ne veut pas dire que les corbeaux soient des agents rationnels qui calculent la hauteur à laquelle s'envoler (H) en fonction de la masse des bulots (M) et de la résistance de leur coquille (R), mais que la relation entre H, M et R est la plus rationnelle ou la plus efficiente pour casser les bulots avec le moindre effort. Peu importe qu'ils effectuent le calcul mentalement, ou que celui-ci leur soit dicté par Dieu, ou qu'il résulte d'un apprentissage par essais et erreurs, ou qu'il soit le produit de l'évolution par sélection naturelle, car quelle que soit la cause du comportement, l'explication téléologique (en termes d'optimalité) reste la même. On peut toujours compléter cette explication téléologique par une explication causale, mais tandis que la première est applicable dans les mêmes termes à d'autres animaux, comme le Gypaète barbu qui lâche des os pour les briser sur les rochers, la seconde peut-être différente dans chacun des cas.

Pour donner un exemple d'optimisation structurelle (les précédents étant comportementaux), nous empruntons celui cité par Huneman (2013c, p. 113-4) à propos du *design* des fibres nerveuses :

> « the first and overriding respect in which the structure of peripheral nerves has been optimized lies in digitalization, ensuring that the information conveyed depends on the pattern and number of impulses transmitted by each fibre, and it is not at the mercy of conduction time that might vary from time to time with local conditions. The second respect is that the size of myelinisation of fibres is closely adapted to their specific function so that the largest ones are preserved for pathways where high speed of conduction is essential, and the smallest and slowest ones are used for sensory pathways where rapidity is not a primary requirement, of for control

> of the autonomous nervous system » (R. Keynes, «The design of peripheral nerves fibers », in *Principles of animal design* [Weibel, Taylor, Bolis], Cambridge University Press, 1998)

Pour une revue de la littérature et des exemples d'application des principes d'optimalité en biologie, voir Parker & Maynard Smith (1990) et Rosen (2013). Voir aussi McNamara (2001) et Edwards (2007).

D'après Oster et Wilson (1978, Chapitre 8) deux des sources d'inspiration des modèles d'optimalité en biologie sont le développement des lois du mouvement en physique classique au XIX^e^ s., d'une part, et l'ingénierie et le design industriel, de l'autre, c'est-à-dire celles auxquelles nous avons eu recours dans ce travail. La troisième est le domaine de l'économie. Les modèles d'optimalité, expliquent, ont une valeur heuristique très importante pour la biologie, mais il convient de ne pas les confondre avec des lois de la nature, et de bien définir la notion d'optimisation dans un cadre mathématique précis :

> « Thus, optimization models are a method for organizing empirical evidence, making educated guesses as to how evolution might have proceeded, and suggesting avenues for further empirical research. At the same time, the concept of optimization can be given a precise definition only in mathematical language. Therefore, it might be a good idea to apply the term only to mathematical models and not to the real world phenomena. Otherwise, like the notions of "instinct" and "drive" in ethology, the concept can create more problems than it solves. » (Oster & Wilson, 1978, p. 295)

De même, d'après John Beatty (1980), ces modèles cadrent mal avec la conception traditionnelle des théories scientifiques, car ne décrivent pas des lois de la nature et doivent plutôt être conçus comme des spécifications de systèmes idéalisés pouvant être employés pour représenter des systèmes empiriques. Il propose de les interpréter plutôt dans le cadre de la conception sémantique des théories scientifiques.

L'un des exemples cités par Michael Ruse (1981) illustre le lien entre les explications fonctionnelles et l'optimalité en biologie : les paléontologistes se sont longtemps demandé quelle était « la fin » ou « l'objectif » ou « la fonction » des plaques osseuses que les stégosaures portaient sur le dos, ou encore quel était le « problème » qu'elles étaient censées « résoudre », et leurs réponses consistaient à dire que ces plaques existaient « pour » faciliter la reconnaissance sociale ou « pour » aider à réguler la température… C'est en raisonnant en ces termes téléologiques, dit Ruse, que les biologistes ont fait de grands progrès dans leur discipline et ont apparemment trouvé la réponse posée par les os du stégosaure, à savoir la thermorégulation. Or, cette conclusion s'appuie sur le *design* de ces plaques osseuses qui les rendait particulièrement adéquates à cette fin, c'est-à-dire sur un raisonnement téléologique en termes d'optimalité. Ce type de raisonnement, poursuit Ruse, se rencontre partout en biologie

évolutionniste où les traits organiques sont analysés et expliqués en termes de problèmes à résoudre et en termes d'optimisation de variables.

Ce que nous avons voulu montrer dans cette section, c'est que les principes et les modèles d'optimalité peuvent être employés dans les sciences, aussi bien en physique qu'en biologie et n'importe où ailleurs, dans la mesure où ils constituent un outil pertinent et utile au chercheur, aussi bien en termes heuristiques qu'explicatifs, et qu'ils sont complémentaires d'autres principes et modes d'explication, sans que cela n'implique un engagement métaphysique d'aucune sorte.

Chapitre X

Comment la téléologie peut-elle être scientifique ? (II)

Dans le chapitre précédent, nous avons mis en parallèle les explications physiques et téléologiques pour chercher à mieux comprendre le fonctionnement de ces dernières et pour réfuter certaines des critiques les plus fréquentes. Nous avons essayé de montrer que les explications téléologiques basées sur le principe d'action efficiente sont analogues aux explications physiques basées sur des principes d'optimalité.

On pourrait penser que les principales différences entre les unes et les autres sont justifiables à partir de la différence entre les objets habituels de la physique classique, d'une part, comme la lumière, les gaz ou les solides inertes, et les objets de la téléologie, de l'autre, c'est-à-dire ceux auxquels on peut attribuer une fin comme les êtres vivants. Depuis cette perspective, les agents téléologiques sont conçus comme des systèmes physiques particuliers qui se distinguent par leur complexité intrinsèque et par leur organisation.

Pourtant, nous avons vu à plusieurs reprises dans ce travail que le critère de complexité n'est pas nécessaire ni suffisant pour les distinguer. Il n'est pas nécessaire parce que l'on attribue des buts et des fins au comportement d'objets sans complexité interne comme le petit cercle de l'expérience de Gergely et Csibra. Et il n'est pas suffisant parce que l'on peut concevoir des systèmes arbitrairement complexes qui ne sont pas pour autant reconnus comme dirigés vers une fin, à l'instar du super-pendule de Bedau (1992a). Les êtres vivants sont à n'en pas douter des systèmes très complexes et cette complexité est ce qui permet d'expliquer que leur comportement soit dirigé vers une fin, mais s'il est dirigé vers une fin ce n'est pas parce que les êtres vivants sont complexes, et ce n'est pas parce qu'ils sont complexes qu'ils sont vivants. Il faut distinguer ce qui fait qu'un système soit considéré comme téléologique et l'explication que l'on donne de ce fait.

Dans les prochaines sections, nous allons tâcher de justifier les explications téléologiques et téléofonctionnelles de façon générale en soulignant ce qui les distingue des explications physiques et en montrant que rien ne s'oppose à leur acceptabilité scientifique.

1. Le pourquoi téléologique ne dépend pas du comment physique

Il y a deux types de systèmes téléologiques : ceux dont le comportement semble dirigé vers une fin, comme le petit cercle et le sauveteur sur la plage, et ceux dont l'existence semble avoir une raison d'être, comme le mécanisme d'Anticythère et les traits biologiques. En ce qui concerne les premiers, si la théorie de Gergely et Csibra est correcte, alors un comportement est dirigé vers une fin quand on peut l'interpréter comme tel sur la base du principe d'action efficiente et pas sur la base d'une définition ou d'une liste de critères. Et ce même principe est responsable à la fois de la catégorisation des agents et de l'identification de la fin vers laquelle tend leur comportement. En ce qui concerne les seconds, ce n'est pas non plus leur organisation ni leur complexité intrinsèque qui nous poussent à leur attribuer une raison d'être comme pourrait le laisser penser l'argument de la montre de William Paley, car nous le faisons aussi pour les outils lithiques de la préhistoire ou pour de simples lignes gravées dans la pierre ou sur des coquillages par nos ancêtres pré-humains (Henshilwood et al., 2002; Joordens et al., 2014). Dans un cas comme dans l'autre, on se tourne vers une explication téléologique lorsque les explications physiques sont possibles mais peu plausibles, ou plutôt lorsque les explications physiques sont à elles seules insuffisantes pour rendre compte du comportement ou de l'existence de la chose en question.

Dans l'expérience de Gergely et Csibra, l'agent n'est autre qu'un petit cercle en deux dimensions qui se déplace sur un écran, sans intériorité ni organisation et sans mécanismes ni propriétés causalement efficaces ; il n'y a donc en principe aucune explication physique à attendre de son comportement. S'il semble bien avoir une fin, il n'a pas de cause[166]. En réponse à la question : « Pourquoi se déplace-t-il comme il le fait ? », on peut expliquer le *pourquoi*, à savoir « *pour* s'approcher du

166 Certes, on peut supposer que son mouvement est déterminé par un programme, auquel cas le comment ferait référence aux algorithmes employés par le programmeur et le pourquoi à ses intentions. Mais si l'on se met à la place des enfants ou si l'on considère que le programme se veut transparent, c'est-à-dire que le comportement de l'agent ne doit pas apparaître comme un objet informatique, alors l'explication téléologique est la seule possible. Et renoncer à ce genre d'explications nous condamnerait à contempler les mouvements du petit cercle sans les comprendre ni les anticiper.

grand cercle », mais on ne peut pas expliquer le *comment*. C'est donc un agent téléologique pur.

Dans le monde réel, les agents sont aussi des systèmes physiques, de sorte que l'on peut formuler à la fois les deux réponses : « *pour*… » et « *parce que*… », mais la première ne dépend pas de la seconde. En effet, un missile autoguidé se déplace comme il le fait « *pour* atteindre sa cible » et « *à cause* du mécanisme *x* », mais l'explication téléologique serait la même si le comportement du missile était causé par les mécanismes *w*, *y* ou *z*. Dans le domaine biologique, certains organismes émettent de la lumière pour se défendre, attaquer ou communiquer, mais la fonction de ce comportement est indépendante du mécanisme causal qui en est responsable et qui varie d'une espèce à l'autre. On peut expliquer pourquoi un organisme le fait dans une situation donnée, c'est-à-dire comment il l'utilise pour attirer ses proies ou repousser ses prédateurs ou pour d'autres fins, tout en ignorant comment il le fait, c'est-à-dire sans savoir qu'il s'agit d'une réaction chimique ni connaître la nature des réactifs impliqués. Inversement, on peut savoir quelles sont les molécules responsables de la bioluminescence tout en ignorant les fonctions qu'elle remplit et dans quelles situations l'organisme la produit. On peut aussi connaître la fonction, mais pas les origines, ou bien les origines, mais pas les fonctions ni le comportement. Sachant par ailleurs que la bioluminescence est apparue plusieurs fois de façon indépendante (voir Haddock, Moline, & Case, 2010), on peut expliquer le pourquoi (la fonction de la production de lumière, la finalité du comportement) indépendamment du comment distal (l'origine évolutive) et du comment proximal (le mécanisme ou la réaction chimique qui en sont causalement responsables).

Considérons maintenant une situation différente. Nous jouons aux échecs et notre adversaire avance un pion. Pour interpréter ce coup, les propriétés physiques du pion (matériau, masse, forme, etc.) n'ont aucune importance, pas plus que l'explication biomécanique du mouvement (muscles, tendons, articulations) ni son explication nerveuse (perception visuelle, activité cérébrale). On pourrait penser que la seule chose qui compte, ce sont les intentions du joueur : « s'il avance son pion, c'est pour protéger son cheval ». Et on pourrait compléter l'explication en disant : « il *sait* que le cheval est menacé par le fou, il *croit* que nous allons l'attaquer et il *désire* le défendre ». Nous lui attribuons ces états mentaux parce que notre adversaire est humain et parce que nous pouvons nous mettre à sa place ou simuler son activité mentale. Toutefois, si notre adversaire n'était pas un humain, mais un extraterrestre déguisé, un poulpe humanoïde, un androïde mécanique ou l'archange Gabriel en personne, cela ne changerait pas notre interprétation de son coup selon laquelle le pion avance pour protéger le cheval. Le pourquoi reste le même, le comment varie.

Un humain est censé avoir des états mentaux intentionnels qui sont censés dépendre causalement de son système nerveux. L'extraterrestre et le poulpe, eux aussi, pourraient avoir des états mentaux, mais avec une « implémentation matérielle » très différente de la nôtre. L'androïde, quant à lui, pourrait fort bien jouer machinalement à partir d'une base de coups préenregistrée. Et pour ce qui est de l'archange, Dieu seul sait le comment du pourquoi. L'identité de notre adversaire et les causes qui le poussent à déplacer le pion n'ont donc aucune importance dans la mesure où elles n'affectent en rien l'explication de son mouvement. Et ses états mentaux intentionnels n'ont pas non plus d'importance, car nous adoptons l'attitude intentionnelle aussi bien face à un humain que face à une machine ou à n'importe quelle autre entité dont le comportement admet cette interprétation (Dennett, 1987).

Il se trouve que nous n'avons pas besoin d'interpréter le mouvement du pion de manière intentionnelle — bien que ce soit peut-être effectivement ce que nous faisons dans la plupart des cas. Nous pouvons en effet nous contenter de dire, en termes purement téléologiques, que « le pion avance pour protéger le cheval », sans implications mentalistes ni références au joueur adverse. L'existence de ce dernier n'est d'ailleurs pas nécessaire, puisque c'est la disposition des pièces sur l'échiquier qui détermine leurs relations respectives de protection, de menace, etc. ; et, parmi les multiples fins possibles d'un même mouvement, certaines sont objectivement meilleures que d'autres, car leur valeur tactique est plus grande[167]. On peut donc déterminer les fins et les moyens dans une situation échiquéenne donnée en appliquant le principe d'action rationnelle ou efficiente : si parmi toutes les conséquences du mouvement du pion la protection du cheval est la meilleure, étant donné la disposition générale des pièces sur l'échiquier et la possibilité d'en déplacer d'autres, alors c'est vraisemblablement pour cela que le pion avance. Il est possible d'envisager, conformément au même principe, plusieurs explications possibles, soit alternatives, soit complémentaires, donnant lieu à des prédictions différentes que l'on peut soumettre au tribunal de l'expérience. Par exemple, prenons le cheval avec le fou et voyons si le pion prend le fou en échange ; s'il ne le fait pas, alors l'explication n'était pas bonne[168].

167 Nous ne prétendons pas nier la dimension psychologique du jeu d'échecs dont on sait qu'elle a joué un rôle très important pour des champions de la taille de Gari Kaspárov et d'Emmanuel Lasker, mais la psychologie est en sus. Par exemple, un « mauvais » coup peut être décisif pour gagner la partie s'il réussit à déstabiliser psychologiquement l'adversaire (Markushin, 2010) (Markushin, 2010), mais si l'on peut parler de « bons » et de « mauvais » coups, c'est d'abord en relation au jeu lui-même, c'est-à-dire à la disposition des pièces sur l'échiquier. Il y a donc une valeur objective des pièces, des mouvements et des configurations, à laquelle s'ajoute une valeur subjective, psychologique.

Ce qui fait qu'un système soit considéré comme téléologique, c'est donc d'abord la possibilité d'interpréter son comportement comme dirigé vers une fin conformément au principe d'action efficiente. Ce comportement est ensuite explicable en termes intentionnels ou causaux selon la nature du système, sa complexité intrinsèque, son organisation, etc., mais ce n'est pas cette explication ni la nature du système qui déterminent son appartenance à la catégorie des agents. C'est le même type de conclusion que celle à laquelle nous étions parvenus au CHAP.V à propos des êtres vivants. Par ailleurs, un même comportement est réalisable par des systèmes très différents, de sorte que l'explication intentionnelle ou causale qui est valable pour les uns ne vaut pas pour les autres. Son explication téléologique, en revanche, ne dépend pas de la nature du système qui le réalise ; elle est plus générale. En ce sens, le pourquoi ne dépend pas du comment. C'est ce que nous défendions plus haut en affirmant que les explications téléologiques ne mentionnent pas les causes.

2. Toutes les explications téléologiques ne se valent pas

Pierre prend son vélo pour aller au travail. Le chat court pour attraper la souris. La luciole s'illumine pour signaler sa présence. Le petit cercle saute pour s'approcher du grand. Le pion avance pour protéger le cheval. Les pierres tombent pour rejoindre leur lieu naturel. Il pleut pour arroser les plantes. Le Soleil brille pour nous éclairer. Les comètes apparaissent dans le ciel pour présager l'avenir. La terre tremble pour punir les péchés.

Toutes les explications téléologiques ne se valent pas. Certaines sont manifestement fausses ou absurdes ou liées à des théories caduques ; d'autres sont vraies ou tenues pour telles. Afin de justifier leur usage dans un cadre scientifique, nous devons obligatoirement faire la part des unes et des autres. Or, suivant la théorie de l'attitude téléologique, c'est le principe d'action efficiente qui justifie leur usage intuitif. La question est de savoir si l'application de ce principe est suffisante pour distinguer le vrai du faux ou les explications valables de celles qui ne le sont pas.

Prenons l'exemple du Soleil. De nombreux enfants et quelques adultes sont convaincus que cet astre brille pour nous éclairer et nous réchauffer, pour rendre la vie possible sur Terre, etc. Pourtant, il ne

168 Lors du second championnat du monde d'échecs des ordinateurs, en 1977, le programme russe Kaissa sacrifia une tour sans raison apparente et perdit la partie peu après. Pour comprendre son comportement, les programmeurs testèrent d'autres options et découvrirent avec surprise que ce coup était le seul qui parait un mat forcé qu'aucun des spectateurs présents n'avait vu.

semble pas que ce soit le moyen le plus efficient pour atteindre ces fins quand on considère que la lumière qui parvient jusqu'à nous n'est qu'une infime portion de toute celle produite, laquelle se perd irrémédiablement dans l'espace. Une étoile plus petite et plus proche de nous serait donc plus efficiente de ce point de vue, mais même dans ces conditions le bilan énergétique continuerait d'être défavorable.

De plus, considérant que le Soleil est une étoile parmi d'autres, qu'il y en a entre 150 et 400 milliards dans la Voie Lactée, et qu'il y a plus de 130 milliards de galaxies dans l'univers exploré (Elbaz, 2007; Siegel, 2015), alors l'explication téléologique ne serait raisonnable que si toutes ces étoiles ou un grand nombre d'entre elles servaient également à illuminer et à réchauffer des planètes habitées. Sinon ce serait un gâchis astronomique. Or, on sait que le nombre de planètes dans notre galaxie est 1,6 fois supérieur à celui des étoiles (Cassan et al., 2012) et que le nombre de corps potentiellement habitables pourrait s'élever à 60 milliards (Yang, Cowan, & Abbot, 2013). Par conséquent, même si toutes les planètes habitables étaient réellement habitées, la proportion d'étoiles qui brillent pour rien serait quatre ou cinq fois supérieure — dans le meilleur des cas — à celle des étoiles qui le font pour éclairer quelque chose ou quelqu'un. L'explication pourrait néanmoins être vraie si le Soleil avait effectivement été créé dans ce but par un dieu, mais elle est difficile à justifier à partir du principe d'action efficiente, car ce dieu-là ne ressemble guère à celui de Fermat et de Maupertuis[169].

En outre, cette explication nous pousse à prédire que le Soleil entendu comme un agent téléologique devrait adapter son comportement aux variations de la situation. Par exemple, si la distance Terre-Soleil augmentait, il faudrait que ce dernier s'approche de notre planète ou augmente son rayonnement, et si l'humanité ou la vie dans son ensemble venaient à disparaître, il devrait l'interrompre ou le réduire. Ces prédictions sont totalement invraisemblables, mais on se souvient que l'un des critères de scientificité d'une théorie selon Karl Popper est justement le fait qu'elle formule des prédictions falsifiables. Or, toutes les données empiriques qui montrent l'existence d'une corrélation entre les variations du rayonnement solaire et le devenir ou l'activité des êtres vivants indiquent que ce sont ces derniers qui s'adaptent à la quantité de lumière reçue et que les émissions solaires sont indifférentes aux réalités

169 Une autre explication possible serait que le dieu en question ait semé l'univers d'étoiles et de planètes comme un agriculteur qui sème de graines un champ labouré, c'est-à-dire en sachant qu'un grand nombre d'entre elles ne fructifieront pas. L'explication est analogue à celle que l'on pourrait donner à propos de la disproportion entre le nombre colossal d'œufs, de graines ou de spores produits par certains êtres vivants et le nombre minuscules de descendants qui en résultent finalement. L'un des inconvénients de cette explication est que si ce dieu est capable de créer des étoiles et des planètes, pourquoi ne les crée-t-il pas de telle façon qu'elles soient toutes fertiles ?

terrestres. Les connaissances théoriques indiquent quant à elles que ces prédictions sont impossibles ou dénuées de fondement. Si elles venaient à se produire, ce serait à n'en pas douter un argument de poids en faveur des thèses créationnistes et finalistes, mais en attendant nous n'avons aucune raison de leur accorder du crédit.

Le principe d'action efficiente implique que les moyens pour atteindre une fin dépendent des contraintes de la situation et varient pour s'adapter à leurs variations, de sorte qu'ils soient toujours les plus efficients possibles. Si nous faisons varier les émissions du Soleil en fonction des conditions sur Terre, nous entrons en contradiction avec les connaissances scientifiques actuelles. Si nous refusons de les faire varier, nous contrevenons au principe d'action efficiente. Dans le premier cas, l'explication téléologique est intéressante (au sens où elle apporte quelque chose que les explications causales n'apportent pas), mais fausse. Dans le second, elle est irréfutable, mais vide, arbitraire et gratuite.

Nous n'allons pas multiplier les exemples. Celui-ci était facile à résoudre et il existe assurément d'autres situations plus compliquées, mais nous ne pouvons pas traiter ici de manière extensive des conditions de validité des explications téléologiques. Il nous suffit d'avoir montré qu'une réflexion sommaire s'appuyant sur quelques données concrètes permet d'identifier une explication invalide dans un cas simple.

Par ailleurs, il serait faux de croire que les explications sans fondement sont l'apanage de la téléologie. Si nous dirigeons notre attention vers la pensée magique, la superstition et les parasciences, nous y verrons énormément d'explications causales fausses, invalides ou infondées comme l'influence des planètes et des étoiles au moment de la naissance sur le caractère et le devenir de la personne, l'idée que briser un miroir cause sept ans de malheur, l'idée que fermer les yeux en faisant l'amour évite la grossesse ou encore les diverses formes par lesquelles la pensée et les paroles agissent sur le monde matériel et les événements distants. Ces explications n'ont rien à envier, de ce point de vue, à l'intervention d'agents surnaturels pour expliquer en termes téléologiques et intentionnels des phénomènes naturels comme le Soleil, la pluie, le tonnerre, les phases de la Lune, les éclipses, les comètes et les tremblements de terre. On ne peut pas jeter l'opprobre sur la pensée téléologique dans son ensemble à cause de ses usages incorrects, pas plus qu'on ne peut rejeter la pensée causale à cause de l'astrologie, de l'alchimie et des croyances superstitieuses en général, ni la méthode scientifique à cause de théories comme celle du phlogistique, du calorique, de l'éther luminifère ou de l'expansion terrestre.

3. La téléologie n'est-elle qu'une illusion cognitive ?

Nous avons défendu au CHAP. VIII que la pensée téléologique correspond à un mode inné d'interprétation du réel, au même titre que la pensée causale ; c'est-à-dire que nous sommes en quelque sorte « programmés » pour penser de cette manière. Or, s'il est vrai que cela rend compte de notre prédilection pour les explications téléologiques, cela ne les justifie pas pour autant. En effet, nous avons aussi une tendance innée à voir des formes et en particulier des visages là où il n'y en a pas. Immédiatement après la naissance, les bébés discriminent et réagissent à des stimuli ressemblant à des visages (Goren et al., 1975) et, plus tard, les enfants et les adultes continuent de voir des formes connues un peu partout : dans la Lune, les nuages, le marc de café, l'écorce des arbres, etc. C'est le phénomène de la paréidolie, une illusion visuelle qui résulte du fonctionnement normal du cerveau (et des logiciels de reconnaissance d'images). Or, si la perception d'un visage n'implique pas nécessairement son existence réelle, la perception de fins et de fonctions chez les êtres vivants n'implique pas non plus qu'ils en aient réellement. Par conséquent, on pourrait penser que l'usage de concepts et d'explications téléologiques en biologie est le fruit d'une illusion cognitive.

Pour répondre à cette objection, il faut d'abord signaler que la paréidolie n'invalide pas notre capacité à reconnaître des visages et des formes ni ne remet en question leur existence. Bien au contraire, nous sommes entourés de personnes et de choses aux contours définis que nous identifions correctement au premier regard grâce à un système de reconnaissance visuelle très effectif. Des erreurs et des illusions se produisent en effet, mais elles sont l'exception, pas la norme, et il est en général assez facile de s'en rendre compte. De la même manière, les explications téléofonctionnelles erronées n'invalident pas le raisonnement téléologique en tant que système interprétatif. Ce qu'elles montrent, c'est que l'attribution de fins et de fonctions n'est pas en soi suffisante pour justifier leur existence, de même que la perception d'un visage sur la Lune n'implique pas qu'il y existe vraiment. Dans un cas comme dans l'autre, on doit donc pouvoir confirmer ou infirmer l'attribution par des moyens indépendants. Pour ce qui est de la Lune, nous n'avons pas besoin d'aller là-haut ni de l'examiner au télescope, car le simple exercice de la raison agrémenté de quelques connaissances scientifiques sommaires devrait être nécessaire pour contredire l'expérience perceptive. Pour ce qui est des explications téléologiques, nous avons vu plus haut avec l'exemple du Soleil comment la réflexion et quelques données concrètes conduisent au même résultat.

On peut ensuite répondre à l'objection en examinant la pensée causale depuis la même perspective, car elle est elle-aussi sujette à des erreurs et des illusions (Bortolotti, 2010). La pensée magique, la superstí-

tion et les parasciences que nous venons de mentionner fourmillent d'exemples d'attributions causales erronées comme la tendance qu'on beaucoup de gens à associer des événements heureux ou malheureux avec le fait de porter certains vêtements ou d'autres objets, de réaliser certains gestes, de prononcer certaines paroles[170].

Une autre illusion commune est celle de la profondeur explicative qui consiste à croire que nous comprenons des phénomènes complexes avec davantage de précision, de cohérence et de profondeur que ce qui est effectivement le cas ; et cette illusion est plus forte avec la connaissance explicative qu'avec d'autres formes de connaissance (C. M. Mills & Keil, 2004; Rozenblit & Keil, 2002). Elle alimente l'extrémisme politique (Fernbach, Rogers, Fox, & Sloman, 2013) et conduit à désigner des boucs-émissaires qui sont considérés comme causalement responsables de ce qui arrive : migrants, homosexuels, juifs. D'après certains auteurs (voir Sloman & Fernbach, 2008), cette tendance à commettre des erreurs systématiques et à succomber à des illusions causales serait justifiée par les bénéfices psychologiques que cela rapporte (aux uns au détriment des autres) : confiance en soi, sentiment de contrôle, réaffirmation des convictions personnelles et collectives, renforcement du lien social.

Indépendamment de ces bénéfices, nous avons aussi fortement tendance à inférer une structure causale entre plusieurs événements à partir de leurs relations temporelles (concomitance, succession), ce qui nous amène à confondre la cause et l'effet lorsque l'ordre temporel est inversé, comme lorsque la jauge de l'essence arrive à zéro avant que la voiture ne s'arrête, à croire qu'un événement est la cause d'un autre alors qu'ils ont une cause commune, ou encore à voir des causes et des conséquences là où il n'y a que des coïncidences ou de simples corrélations (Burns & McCormack, 2009; Lagnado & Sloman, 2006).

La confusion entre corrélation et causalité est d'ailleurs tellement répandue qu'on la rencontre dans des revues scientifiques sérieuses comme cet article du *New England Journal of Medicine* qui s'appuie sur une corrélation statistique pour établir (sérieusement) un lien causal entre la consommation de chocolat et l'obtention du Prix Nobel (Messerli, 2012), secondé par un article dans *Nature* qui confirmait que les lauréats

170 Björn Borg semblait croire que sa barbe avait quelque chose à voir avec son succès au tournoi de Wimbledon et Michael Jordan a porté pendant toute sa carrière le short de son équipe universitaire d'origine en dessous de son short officiel. S'il y a une quelconque relation causale entre les victoires de ces sportifs et leurs manies, elle tient vraisemblablement au fait que les rituels, les routines et les interdits sont psychologiquement rassurants et qu'ils donnent une illusion de contrôle. Une relation qui n'est apparemment pas propre à l'humain puisque Konrad Lorenz rapportait, dans un de ses livres, que certains animaux ont eux-aussi des rituels apparemment absurdes qui les aident à réduire leur anxiété.

sont de grands mangeurs de chocolat (Golomb, 2013). Une corrélation comparable à celles entre la consommation de mozzarella et l'obtention d'un doctorat en ingénierie civile, entre la consommation de margarine et le taux de divorce dans l'État de Maine, entre les apparitions de Nicolas Cage au cinéma et les morts par noyade à la piscine[171] ou encore entre les populations locales de cigognes et les taux de naissance de bébés humains (Matthews, 2000).

Les détracteurs de la téléologie, comme David Hanke (2004) et Paul Kramer (1998), l'accusent d'être une source d'erreurs, de confusions et de mauvaise science, et ils estiment pour cette raison qu'elle devrait être mise au banc de la science. En appliquant le même critère, il faudrait aussi proscrire les statistiques. Et si les explications téléologiques sont parfois le fruit d'une illusion cognitive, on peut en dire autant des explications causales. Cela n'implique pas que les premières soient justifiées, mais qu'on ne peut pas les rejeter en bloc, ou du moins pas pour cette raison.

4. Aristote, Newton, Einstein et le problème de l'ontologie des sciences

Une autre bonne raison de rejeter la téléologie depuis une posture naturaliste est de considérer que les fins et les fonctions n'existent pas, qu'elles ne font pas partie de l'ontologie de la nature ni de celle des sciences. Nous n'allons pas nous engager ici dans une discussion ontologique qui dépasse largement le cadre de ce travail. Une simple réflexion sur la nature des fins ou sur la vérité des attributions téléologiques pourrait occuper plusieurs chapitres, voire plusieurs thèses. Pour répondre à la critique, nous allons donc nous contenter d'illustrer le problème à partir d'un exemple historique concret : le dépassement de la physique d'Aristote par celle de Newton.

Il ne fait aucun doute que la physique newtonienne est une avancée sans précédent dans la compréhension de l'univers et de ses lois. Entre autres choses, elle marque l'abandon définitif de l'étude qualitative des phénomènes naturels au profit d'une approche mathématique ainsi que la disparition des causes finales au profit des causes efficientes. Si les deux théories n'étaient pas incommensurables (Kuhn, 1957, 1983), on dirait que du point de vue de Newton les explications téléologiques d'Aristote sont non seulement innécessaires, mais fausses. Les corps lourds ne tombent pas pour rejoindre leur lieu naturel, car il n'y a pas de lieux naturels, pas plus qu'il n'y a de mouvements naturel et violent. Pour le physicien anglais, les mouvements obéissent à des lois universelles qui emploient la notion de force entendue au sens de cause effective. S'ils

171 Ces corrélations existent bel et bien ! Voir le site web : « Spurious Correlations » (www.tylervigen.com).

tombent, ce n'est pas en vertu d'une tendance naturelle inhérente qui les pousse vers le centre de l'univers, mais à cause de la gravitation. Par conséquent, le dépassement d'Aristote par Newton semble montrer que le progrès scientifique passe par l'abandon des explications téléologiques, puisque les changements dans la nature n'obéissent pas à des fins ni à des tendances intérieures, mais à des lois causales. Pour cette même raison, pensent certains, la biologie devrait abandonner elle-aussi les fonctions et les fins pour devenir une science mûre.

En 1915, Albert Einstein publia sa théorie de la Relativité générale que beaucoup considèrent comme le plus grand accomplissement de la pensée scientifique depuis les *Principia Mathematica* de Newton en 1687. Or, cette théorie explique le phénomène de la chute des corps sans recourir à la gravitation. Einstein abandonne en effet l'espace-temps uniforme et homogène de Newton au profit d'un espace-temps dont la courbure est variable et dont la variation dépend de la matière ; une variation de courbure qui explique les mêmes phénomènes qu'expliquait précédemment la gravitation. Dès lors, en termes ontologiques, selon cette théorie, la force gravitationnelle n'existe pas !

Newton lui-même était conscient du caractère mystérieux de cette force attractive universelle d'origine inconnue agissant à distance et de manière instantanée, ce qui contredisait les conceptions mécanistes de son époque. Officiellement, il ne se prononçait pas sur l'interprétation ontologique de la gravitation— hormis le recours à l'action directe de Dieu pour contrebalancer son action et assurer la stabilité de l'univers — et se limitait à l'analyse mathématique des phénomènes. Officieusement, son intérêt pour l'alchimie et les phénomènes non mécaniques l'amenait à admettre qu'il existe dans la nature des forces incompréhensibles. Jusqu'à la fin du XIX^e^ s., certains physiciens de renom comme Huygens, Helmholtz et Hertz se montrèrent méfiants à l'égard de cette force, au point de proposer son exclusion des notions fondamentales de la mécanique. Pour d'autres, au contraire, elle servit de modèle de référence pour la description d'autres phénomènes non mécaniques comme les interactions électromagnétiques. On peut donc dire que la gravitation était comme une maîtresse dont aucun physicien ne pouvait se passer, mais en compagnie de laquelle certains préféraient ne pas être vus en public.

Aujourd'hui, la situation a changé. Nous vivons dans un monde relativiste et quantique où la notion de force n'a plus beaucoup de sens. Pourtant, elle demeure indispensable. Non seulement les physiciens parlent constamment de forces pour analyser les phénomènes, mais ils continuent d'employer les équations de Newton au lieu de celles d'Einstein dans un grand nombre de cas. La gravitation n'a pas disparu des manuels scolaires, ni universitaires, ni des publications scientifiques, ni des applications pratiques comme le lancement d'objets dans l'espace par la NASA. Inutile en effet de recourir à des concepts physiques et des outils

mathématiques relativistes lorsque les concepts et le formalisme newtoniens, beaucoup plus intuitifs et plus simples, font parfaitement l'affaire. Loin d'être une manifestation de schizophrénie ontologique, c'est au contraire un exemple de pragmatisme scientifique. Il serait absurde de vouloir se priver d'une approche qui continue de porter ses fruits sous prétexte que les équations de Newton sont fausses, que l'analyse des forces est dangereuse ou que la gravitation n'existe pas.

La conclusion que nous tirons de cette petite histoire, c'est qu'il faudrait peut-être se montrer plus prudent, peut-être plus flexible, lorsqu'il s'agit d'autoriser ou d'interdire l'entrée de nouvelles entités dans notre ontologie scientifique, et aussi plus tolérant à l'égard de formes d'explication qui sont intuitives, efficaces et utiles, même si elles ne sont pas strictement « vraies » au sens où elles ne correspondent pas à une prétendue réalité extérieure que les sciences auraient la mission de décrire. D'autant plus que la théorie de la vérité comme correspondance est pour le moins discutable et que l'ontologie des sciences, y compris en physique, est un terrain trop mouvant pour qu'on puisse assurer que ce qui est réel aujourd'hui (les particules du Modèle Standard, par exemple) le sera encore demain. Quoi qu'il en soit, nous reviendrons sur la réalité des fins et la vérité des attributions téléologiques lorsque nous aborderons la question de la justification de la téléologie depuis les approches mentaliste, valorative et organisationnelle.

5. Action efficiente et optimalité fonctionnelle

Nous avons jusqu'ici tâché de justifier le recours à la téléologie de manière générale à partir de la pensée téléologique et du principe d'action rationnelle ou efficiente, c'est-à-dire à partir de l'interprétation du comportement des agents. Or, le problème principal de ce travail est la définition des fonctions, lesquelles sont attribuées non seulement à des comportements, mais aussi aux parties d'objets biologiques et techniques. Il nous faut donc montrer que cette justification est également applicable aux fonctions biologiques.

Nous allons avancer ici quelques arguments en ce sens, mais nous devons cependant exprimer une certaine réserve quant à la portée des conclusions que l'on pourrait en tirer. Il nous semble en effet que la défense de la téléologie face aux objections et critiques qu'on lui a opposées est applicable aux fonctions biologiques dans la mesure où le concept de fonction est téléologique. Toutefois, il nous semble aussi que les similitudes et les différences entre les explications téléologiques fonctionnelles et non-fonctionnelles méritent d'être explorées plus en détail. Il en va de même pour les explications fonctionnelles qui portent sur une action ou un comportement et celles qui portent sur une structure ou un processus dans un système.

La notion de comportement fait référence aux actions coordonnées des agents (individus et groupes) en réponse à un stimuli interne ou externe[172]. Les fonctions font elles-aussi référence à des actes et des activités (voir *Trésor de la Langue Française*), ou plus généralement à des opérations (Ferrater Mora, 1986), ou seulement à leurs effets (L. Wright, 1973), ou plutôt à leurs dispositions (Cummins, 1975), étant donné qu'une fonction peut ne jamais être réalisée et demeurer à l'état de réalisable (Arp & Smith, 2008). Il y a donc manifestement une idée commune aux comportements et aux fonctions autour du champ sémantique de l'action et de ses conséquences.

De plus, à la suite de Wimsatt (1972, 2002) et de Gayon (2010b), nous estimons que les fonctions ne devraient pas être attribuées aux parties d'un organisme, mais aux comportements de ces parties, dans des conditions internes et environnementales spécifiées. C'est d'ailleurs en ces termes que nous avons formulé notre définition téléologique des fonctions au CHAP. VII. Depuis cette perspective, le problème de la distinction entre comportements et parties ne se pose pas, et notre justification de la téléologie est également applicable au comportement des agents qu'aux fonctions biologiques ou techniques.

Cela étant, on peut arriver à la même conclusion par des moyens différents. Par exemple à partir de la différence entre les parties et le tout. On attribue habituellement un comportement à un organisme entier ou à un groupe d'individus, c'est-à-dire à un tout, tandis que les fonctions ne sont généralement attribuées qu'à des parties. Cependant, le comportement lui-même a une fin ou une fonction selon qu'on le considère comme un tout ou comme une partie. La parade nuptiale du paon, par exemple, a une finalité lorsqu'elle est considérée en elle-même et une fonction quand on la considère dans le cadre du comportement sexuel dans son ensemble. De même, on peut dire que la reproduction est une fin pour l'individu et qu'elle a une fonction pour l'espèce. On peut aussi décomposer un comportement complexe, comme la parade du paon, et attribuer des fonctions à ses éléments dans la mesure où ils contribuent à la finalité de l'ensemble. À l'inverse, on peut isoler un élément d'un ensemble et lui attribuer un comportement. Par exemple, on peut qualifier de comportement l'activité des cellules du tissu conjonctif (Couchman & Rees, 1979) ou d'un organe comme le cœur (Ritzenberg, Adam, & Cohen, 1984). La différence entre un comportement et une

172 Pour une définition plus précise, voir Levitis *et al.* (2009) : « Behaviour is the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes. »

activité fonctionnelle est donc avant tout une question de point de vue et de relation hiérarchique.

Dans la justification que nous avons faite des explications téléologiques, deux des exemples que nous avons analysés sont celui du jeu d'échecs et celui du Soleil. Quand, au jeu d'échecs, on dit que le pion avance pour protéger le cheval, on considère cette pièce comme un agent et on attribue une finalité à son comportement sur la base du principe d'action efficiente. Cependant, on peut aussi lui attribuer une fonction en considérant qu'il joue un rôle stratégique, par exemple dans la préparation d'une attaque coordonnée du joueur adverse. De manière analogue, nous avons rejetté l'explication téléologique selon laquelle le Soleil brille pour nous éclairer en considérant celui-ci comme un agent et en attribuant à son comportement la finalité en question, mais on pourrait aussi bien le considérer comme un artefact et dire que le fait de briller est sa fonction, comme dans le récit biblique, ou encore la lui attribuer au terme d'une analyse fonctionnelle de la capacité de notre système solaire à abriter la vie, c'est-à-dire en le considérant comme une partie d'un ensemble plus complexe au sein duquel il a un rôle à jouer. Dans un cas comme dans l'autre, qu'on lui attribue un comportement téléologique ou une fonction, cela n'affecte pas notre argument.

Par ailleurs, il existe sans doute une relation psychologique très forte entre l'interprétation des comportements téléologiques et l'attribution de fonctions. (Nous réitérons ici nos réserves, car les études de psychologie cognitive ne sont pas encore très concluantes quant à la nature de cette relation.) Nous avons vu que la pensée téléologique, dont l'origine se trouve sans doute dans le besoin primitif d'anticiper les mouvements des proies et des prédateurs, permet d'interpréter le comportement d'un agent en termes de moyens et de fins en tenant compte des contraintes de la situation. Ces contraintes incluent des objets statiques — comme le rectangle de l'experience de Gergely et Csibra — qui font obstacle au mouvement de l'agent ou qui au contraire le facilitent. La pensée téléologique permet donc d'inclure la présence et les propriétés d'objets statiques dans l'interprétation du comportement d'un agent. Dès lors, ces objets peuvent être considérés comme des moyens. C'est là l'une des sources possibles de notre capacité à utiliser des outils. S'il en est ainsi, c'est-à-dire si la pensée fonctionnelle s'appuie sur la pensée téléologique, alors les fonctions sont essentiellement des relations téléologiques partielles caractérisant le rôle — c'est-à-dire l'action ou les effets de la présence — d'un élément au sein du tout plus complexe auquel il appartient[173].

173 On trouve dans le *Grand dictionnaire de la philosophie* (Blay, 2003) une définition qui se rapproche de cette idée : « [Fonction biologique] Dans la totalité complexe d'un organisme, activité spécifique d'un organe, faite en vue de la structure complète qui en recueille les effets. » Elle rejoint l'idée formu-

De plus, il existe aussi vraisemblablement un rapport entre les principes d'optimalité qui guident l'interprétation des actions téléologiques, d'un côté, et des relations fonctionnelles, de l'autre. Les deux sont des principes fondés sur l'efficience des moyens pour arriver à une fin dans une situation donnée. La différence est que le premier porte sur le comportement des agents et le second sur la constitution des objets. Nonobstant, le principe d'action rationnelle est applicable dans un cas comme dans l'autre. En effet, nous avons montré précédemment qu'il est applicable à l'usage et à la fabrication d'outils chez les animaux non-humains, c'est-à-dire que le choix des propriétés fonctionnelles pertinentes de ce genre d'outils, comme un caillou pour écraser des noix ou une brindille pour débusquer des insectes, est directement lié au principe d'action rationnelle au sens où les animaux sont capables de sélectionner ou de fabriquer les moyens les plus efficients pour atteindre leurs fins, étant donné les contraintes de la situation. En poussant l'argument un peu plus loin, on peut dire que la fabrication d'outils complexes, comme une montre, est liée au même principe. Quand l'horloger sélectionne ou fabrique les éléments qui feront partie de la montre, on peut supposer que son comportement est rationnel et qu'il obéit à un principe d'efficience, c'est-à-dire qu'il choisit les meilleurs moyens à sa disposition pour arriver à ses fins. S'il en était autrement, la rétro-ingénierie ne serait pas possible. Or, celle-ci ne s'intéresse pas aux actions de l'agent, mais aux produits de ses actions, c'est-à-dire aux artefacts et à leur organisation. La différence entre les deux principes d'optimalité est donc essentiellement une question de perspective[174].

lée par Claude Bernard (1865) selon laquelle « les phénomènes physiologiques complexes sont constitués par une série de phénomènes plus simples qui se déterminent les uns les autres en s'associant ou se combinant pour un but final commun ».

174 Pour comprendre les fonctions d'un objet archéologique, il faut adopter à la fois les deux perspectives. D'un côté, nous voulons comprendre le fonctionnement de l'objet. De l'autre, nous voulons connaître les motivations de ceux qui l'ont fabriqué. Mais nous voulons aussi que les deux explications coïncident. On sait que la « batterie de Bagdad » est capable de produire un courant électrique de faible intensité, mais on peut douter que ses créateurs aient eu connaissance de cette propriété et qu'ils en aient vu l'utilité ; on peut donc douter que cet objet ait été créé avec l'intention d'être une batterie. À l'inverse, on peut inférer que le mécansime d'Anticythère est une horloge astronomique parce que, comme dit Paul Bloom (1996, p. 12) : « la meilleure explication que l'on puisse donner de son apparence et de son usage potentiel est qu'il résulte de l'intention de créer un artefact de ce type ». De manière analogue, pour comprendre les fonctions d'un objet biologique, on peut adopter une perspective centrée sur l'objet, centrée sur l'agent, ou les deux à la fois, comme nous avons pu le constater à plusieurs

Chapitre XI

Qu'apporte la téléologie à la biologie ?

Jusqu'ici, nous avons essayé de montrer que la téléologie n'est pas incompatible avec le discours et les connaissances scientifiques et que rien ne s'oppose à ce qu'elle puisse être employée dans un cadre scientifique. Nous voulons maintenant montrer que la téléologie est non seulement acceptable du point de vue scientifique, mais qu'elle est indispensable à la biologie, de sorte que les biologistes ne devraient pas avoir honte d'être vus avec elle en public.

Après avoir rejeté l'accusation d'anthropomorphisme, nous montrerons que la téléologique biologique permet de réaliser des prédictions qui seraient difficiles ou impossibles à réaliser autrement. Nous verrons aussi qu'elle joue un rôle très important dans la classification des êtres vivants, laquelle rend possible de nombreuses inférences. Finalement, nous montrerons que les explications téléologiques et fonctionnelles sont non seulement des explications authentiques, du point de vue de la psychologie cognitive, mais qu'elles sont aussi scientifiquement légitimes, du point de vue de la philosophie des sciences.

1. Le problème de l'anthropomorphisme

La téléologie biologique est souvent accusée d'anthropomorphisme. Est-ce une accusation sérieuse ? D'un côté, nous avons vu dans le chapitre précédent que la téléologie n'est pas nécessairement chargée de connotations anthropomorphiques, par exemple lorsque les physiciens raisonnent sur un système à partir de principes variationnels, ou lorsque les biologistes parlent de l'adaptation d'une espèce d'organismes à des

reprises dans ce travail. En ce qui concerne l'agent, c'est normalement à la sélection naturelle ou au processus évolutif en général qu'est dévolu ce rôle.

contraintes environnementales. De l'autre, il est vrai qu'un certain nombre de biologistes emploient effectivement ce type de discours non seulement dans des textes de vulgarisation mais aussi dans leurs publications scientifiques. La question est de savoir quelles sont ses implications, s'il est lié à la téléologie et dans quelle mesure il porte préjudice à la biologie en tant que discipline scientifique.

Lorsque des biologistes comme Richard Dawkins disent que les gènes sont égoïstes et comparent la sélection naturelle à un horloger aveugle, prennent-ils au sérieux leurs propres affirmations, comme le laissent entendre certains critiques (J. Davies, 2010), ou ne font-ils qu'employer des métaphores pour attirer l'attention des lecteurs, ou encore parce qu'ils estiment que c'est une façon adéquate d'aborder les problèmes (Kramer, 1998) ? On peut difficilement croire que des scientifiques puissent sérieusement attribuer des états mentaux à des gènes ou des bactéries, ni qu'ils prennent au pied de la lettre des expressions et analogies comme « la course aux armements » pour caractériser les relations proie-prédateur, ou l'« arbre de la vie » pour décrire les relations entre les différentes espèces présentes et passées. Quoi qu'il en soit, même quand on ne les prend pas au sérieux, les métaphores ne sont pas neutres. Sans que nous en soyons conscients, elles exercent une influence non négligeable sur les décisions que nous prenons et sur notre façon de penser et de saisir les problèmes, et c'est là que le bât blesse, disent certains, car les métaphores en biologie sont trompeuses (Ball, 2011).

L'anthropomorphisme et l'usage des métaphores et des analogies en général est un problème en tant qu'il véhicule des images séduisantes mais fausses qui peuvent provenir ou se convertir en obstacles épistémologiques, c'est-à-dire en foyers de résistance dans la constitution des concepts et des théories scientifiques (Bachelard, 1938).

C'est un problème qui affecte toutes les sciences et qui n'est pas particulièrement lié à la biologie ni à la téléologie. La théorie de Newton, avec ses notions d'attraction, de répulsion et de force, a été accusée d'anthropomorphisme et d'animisme par ses contemporains, de même que les tubes de force de Faraday — que Maxwell comparait à des muscles — ont été accusés d'anthropomorphisme par les newtoniens (Agassi, 2008, p. 478-479). Comme le remarquait Darwin lui-même, dans un passage où il justifie son usage du concept de sélection naturelle contre cette même accusation, les chimistes[175] emploient la notion d'affinité, issue de

175 « Several writers have misapprehended or objected to the term Natural Selection. […] [They] have objected that the term selection implies conscious choice in the animals which become modified; and it has even been urged that, as plants have no volition, natural selection is not applicable to them! In the literal sense of the word, no doubt, natural selection is a false term; but who ever objected to chemists speaking of the elective affinities of the various elements?—and yet an acid cannot strictly be said to elect the

l'alchimie et définie seulement au XX^e^ s., malgré ses connotations dans le même sens ; ils n'hésitent pas à parler d'orbitales atomiques comme si les électrons étaient des planètes en révolution autour du Soleil, et ils n'hésitent pas non plus à dire par exemple que « l'azote a besoin de trois électrons pour compléter son octet ». La science des matériaux parle quant à elle de métaux à mémoire de forme. La physique fondamentale parle de la durée de vie des particules. Et la notion même de loi, que l'on rencontre dans toutes les sciences, de la physique à la sociologie, n'est pas moins anthropomorphique que celle de stratégie dont Kramer critique l'usage biologique.

Dans son introduction à l'édition du 30^e^ anniversaire du *Gène égoïste* (1976), Dawkins revient longuement sur le titre de l'ouvrage et sur l'anthropomorphisme de son contenu. Il se justifie tout d'abord en disant que la personnification des gènes ne devrait pas poser de problème, car il serait totalement absurde de croire que les molécules d'ADN ont une personnalité, et qu'aucun lecteur sensé ne pourrait accuser l'auteur de le faire. Il continue en montrant que de grand noms de la science pratiquent l'anthropomorphisme. Parmi eux, Jacques Monod aurait affirmé que lorsqu'il réfléchit à un problème de chimie, il se demande ce qu'il ferait s'il était un électron. La personnification n'est donc pas seulement un outil pédagogique, ajoute Dawkins, mais un outil de pensée qui peut réellement aider le chercheur à trouver la réponse correcte à un problème en lui évitant certaines erreurs et certains pièges que peuvent parfois poser les calculs.

D'autres auteurs ont décrit le rôle crucial que jouent les métaphores dans la pratique des scientifiques, non seulement dans leur manière de comprendre le monde qui les entoure et les phénomènes qu'ils étudient, mais aussi dans les explications qu'ils en donnent, aussi bien en biologie que dans d'autres domaines (T. L. Brown, 2003; Keller, 2003).

L'anthropomorphisme est en effet lié au fonctionnement intime de l'esprit et du langage. George Lakoff et Mark Johnson (2003) ont montré que les métaphores structurent notre système conceptuel et notre expérience du monde, qu'elles rendent possibles de nouveaux modes de compréhension de ce que nous faisons et percevons, et qu'elles jouent un rôle central dans la pensée abstraite.

De même, d'après Steven Pinker (2000), si nous employons des métaphores, ce n'est pas tant pour emprunter des mots que pour emprunter leur mécanisme de déduction et raisonner sur d'autres sujets, pour couvrir de nouveaux domaines de connaissance et produire des idées de plus en plus complexes en combinant métaphoriquement des concepts élémentaires. Ce mécanisme cognitif est un aspect fondamental du raisonnement scientifique :

---

base with which it in preference combines. » (Darwin, 1872, Chapitre 4) (Darwin, 1872, Chapitre 4)

> Même le raisonnement scientifique le plus profond est un assemblage de métaphores de la vie quotidienne. Nous libérons nos facultés des domaines pour lesquelles elles étaient conçues et nous utilisons leurs mécanismes pour expliquer des domaines nouveaux qui ressemblent de manière abstraite à des domaines anciens. Les métaphores dans lesquelles nous pensons proviennent non seulement de scénarios de base comme ceux où l'on se déplace et se cogne, mais de modes de connaissance entiers. Pour faire de la biologie à l'université, nous prenons notre mode de compréhension des objets fabriqués et nous l'appliquons aux organismes. Pour faire de la chimie, nous traitons l'essence d'une catégorie naturelle comme une collection de minuscules objets bondissants et collants. Pour faire de la psychologie, nous traitons l'esprit comme une espèce naturelle. (Pinker, 2000, p. 380)

Depuis cette perspective, on devrait se demander si l'anthropomorphisme et l'usage des métaphores en général est véritablement un problème pour les sciences qu'il faudrait éliminer de leurs discours ou si, au contraire, il s'agit d'un aspect essentiel de la pensée scientifique qui présente ses avantages et ses inconvénients, mais que l'on ne peut pas écarter. René Thom disait par exemple que l'usage de la métaphore n'est pas en soi un mal et que, toute connaissance étant métaphorique, il est sans doute inévitable[176].

Richard Lewontin (2001a) affirme quant à lui qu'il semble impossible de faire des sciences sans métaphores et que certaines des découvertes majeures de la biologie, comme celle de la relation entre l'ADN et les protéines, sont directement liées à l'usage de métaphores telles que « programme », « code », « langage » et « information » pour parler du génome — à une époque où la mode était aux briseurs de code et à la théorie de l'information. Il explique par ailleurs que les métaphores sont la seule manière que nous ayons de comprendre humainement des phénomènes qui se situent au-delà de notre expérience ordinaire :

> « It is not possible to do the work of science without using a language that is filled with metaphors. Virtually the entire body of modern science is an attempt to explain phenomena that cannot be experienced directly by human beings, by reference to forces and processes that we cannot perceive directly because they are too small, like molecules, or too vast, like the entire known universe, or the result of forces that our senses cannot detect, like electromagnetism, or the outcome of extremely complex interactions, like the coming into being of an individual organism from its conception as a fertilized egg. Such explanations, if they are to be not merely formal propositions, framed in an invented technical language, but are to

176 René Thom, *Apologie du logos*, Hachette, 1990, p. 641, cité par Jean-Jacques Wunenburger (2000).

> appeal to the understanding of the world that we have gained through ordinary experience, must necessarily involve the use of metaphorical language. » (Lewontin, 2001b, p. 3)

Mais l'usage des métaphores comporte aussi des risques, ajoute cet auteur, car on finit souvent par confondre l'image et la réalité. Cela nous conduit à attribuer des propriétés et à poser des questions à propos de notre objet d'étude qui ne font que renforcer l'image métaphorique originale et à perdre de vue les aspects du système qui ne s'y ajustent pas.

D'après Stephen Jay Gould (1993), Darwin fut un maître de la métaphore et son succès est dû en grande partie à son sens particulier des comparaisons qui facilitent la compréhension, la sienne et celle des lecteurs ; mais ces images, ajoute-t-il, peuvent elles-mêmes devenir des obstacles qui nous empêchent de voir ou de comprendre le processus évolutif depuis une perspective nouvelle et plus fidèle à la réalité.

Il y aurait beaucoup à dire sur cette question, mais ce n'est pas l'objet de ce travail et les quelques considérations précédentes nous semblent suffisantes pour absoudre la téléologie biologique, en tant que telle, du péché d'anthropomorphisme. On ne peut nier que les descriptions et les explications téléologiques se prêtent volontiers à des interprétations anthropomorphiques, mais il ne s'agit pas d'un problème exclusif de la téléologie, ni de la biologie, ni des sciences en général.

Ce n'est d'ailleurs pas un problème en soi, mais un outil mental qui permet aux primates que nous sommes d'appréhender et d'explorer un monde beaucoup plus vaste que celui de notre expérience ordinaire. C'est un outil de pensée qui, comme d'autres, comporte ses risques et ses bénéfices. Depuis cette perspective, l'anthropomorphisme n'est donc pas une raison suffisante pour exclure la téléologie biologique du domaine des sciences.

Cela étant, nous partageons aussi le message de prudence que Lewontin attribue à Rosenblueth et Wiener : *le prix de la métaphore est une vigilance éternelle*.

2. Prédictions téléologiques de l'évolution du vivant

Les prédictions sont l'un des éléments essentiels de la méthode scientifique telle qu'elle a été conçue à partir du modèle de la physique classique, en particulier parce qu'elles permettent de confronter les hypothèses théoriques à l'expérience et à l'expérimentation. Or, il existe une idée assez répandue selon laquelle la biologie, contrairement à la physique, ne formule pas de prédictions. Cette idée est vraisemblablement liée aux écrits de Karl Popper affirmant que le darwinisme n'est pas

une théorie scientifique susceptible d'être testée et falsifiée, mais un programme de recherche métaphysique et tautologique, ou encore que son pouvoir explicatif et prédictif est quasi inexistant (voir par exemple Popper, 2005). Bien que sa pensée ait pu être mal interprétée et qu'il soit lui-même revenu sur ses propos, l'idée est restée.

Pourtant, la biologie formule bel et bien des prédictions. Les paléontologistes avaient par exemple prédit l'existence d'un animal fossile intermédiaire entre les poissons et les tétrapodes ayant vécu au Dévonien dans des cours d'eau douce de faible profondeur. Après plusieurs années de fouilles infructueuses, un fossile aux caractéristiques requises a finalement été découvert (Daeschler, Shubin, & Jenkins, 2006).

Une autre idée commune, que l'on peut faire remonter au moins jusqu'à l'époque de Darwin, est que l'attribution de fins aux processus évolutifs et l'explication téléologique des traits organiques sont fausses, tout simplement. L'erreur consisterait à croire en une finalité intentionnelle (créationnisme, Dessein Intelligent) ou non-intentionnelle (finalisme, vitalisme) qui préexiste et qui guide ou qui détermine le devenir des formes et des fonctions biologiques[177]. Darwin a montré que de telles suppositions ne sont pas plus nécessaires que l'hypothèse de Dieu dans le système du monde de Laplace et les progrès de la discipline ont ensuite confirmé qu'aucune fin préétablie ne dirige l'évolution du vivant. Depuis cette perspective, la critique est donc totalement justifiée, mais elle ne concerne pas le débat dont il est question dans ce travail, car la plupart des biologistes et des philosophes qui emploient et qui défendent aujourd'hui le langage téléologique n'en acceptent pas moins la théorie darwinienne. Par ailleurs, cette critique peut conduire à penser que les prédictions téléologiques de l'évolution du vivant sont fausses ou infondées puisque l'avenir n'est pas écrit.

Il faut cependant distinguer entre la croyance à la prédétermination de l'évolution du vivant et la prédiction des résultats de cette évolution. La première implique une thèse substantive sur le monde ; la seconde est une hypothèse plus ou moins justifiée qui, au moment où on la formule, n'est ni vraie ni fausse et dont la confirmation ou l'infirmation en disent davantage sur celui qui la prononce que sur le monde lui-même.

Par exemple, on aurait pu prédire il y a quelques années l'apparition d'organismes capables de décomposer le plastique pour s'en nourrir, étant donné l'accumulation vertigineuse de ce matériau dans la nature depuis

177 En ce qui concerne la téléologie historique, la croyance peut se décliner en deux versions : forte et faible. La version forte affirme que le l'avenir est écrit et que, par exemple, l'univers devait nécessairement aboutir à l'existence d'un être intelligent et conscient. La version faible affirme qu'il existe une orientation dans l'évolution, par exemple vers une complexification croissante des formes du vivant, et donc une détermination du processus, mais pas nécessairement une prédétermination de l'avenir.

cinquante ans et l'avantage adaptatif que cela leur conférerait. Or, il semble que cette (rétro-)prédiction soit en train de s'accomplir (Bombelli, Howe, & Bertocchini, 2017; Kinoshita, Kageyama, Iba, Yamada, & Okada, 1975; Reisser et al., 2014; Russell et al., 2011; Yang, Yang, Wu, Zhao, & Jiang, 2014). De même, on pouvait prédire depuis longtemps que les pathogènes allaient muter pour résister aux antibiotiques, ce qui est effectivement arrivé ; ou encore que les insectes allaient s'adapter aux plantes génétiquement modifiées pour produire des pesticides[178]. Cela ne veut pas dire que les organismes aient pris la décision, pour leur bien, d'entreprendre leur évolution. Cela ne veut pas dire non plus que Mère Nature ou qu'un agent surnaturel, dans leur infinie sagesse, aient guidé cette évolution. Et cela n'implique pas davantage que l'avenir fût écrit à l'avance ni qu'il devait se produire, ni même que les lois de l'univers conduisaient nécessairement à ce résultat, lequel est contingent. Mais cela donne à penser que le résultat en question n'est pas tout à fait le produit du hasard, bien que l'évolution par sélection naturelle repose sur des mécanismes partiellement aléatoires. Nous nous situons en effet sur un terrain à cheval entre le hasard et la nécessité, comme dirait Monod.

En physique classique, les prédictions s'appuient soit sur un mécanisme causal sous-jacent soit sur une régularité nomique, et la violation des attentes est l'expression d'une ignorance de la part du sujet. Dans le domaine de l'évolution biologique, la violation des attentes est plutôt l'expression d'une certaine contingence ou indétermination de l'avenir et les prédictions des exemples précédents ne résultent pas de l'application de lois causales ou probabilistes, mais de la considération de principes généraux. Car loin d'être en contradiction avec la théorie darwinienne, c'est sur elle que repose la possibilité de telles prédictions.

Il faut ensuite distinguer les prédictions qui s'appuient sur la considération des causes et celles qui partent de l'attribution d'une fin. Par exemple, si je lâche le verre que je tiens dans la main, je peux prédire qu'il va se briser. Pour cela, je tiens compte de la fragilité du verre, de sa masse, de la dureté du sol et de la distance entre les deux. C'est une prédiction mécanique qui part des causes pour en déduire les conséquences. Mais je peux aussi prédire que, si ce n'est pas moi qui le laisse tomber, il finira tout de même par se briser, un jour ou l'autre, quoi qu'il arrive. Pour réaliser cette prédiction, il me suffit de penser que n'importe qui d'autre peut le faire tomber, qu'il peut recevoir un choc lors d'un déménagement, souffrir les effets d'un tremblement de terre ou disparaître avec le reste de la planète lorsque le Soleil en fin de vie l'engloutira. C'est là aussi une prédiction mécanique qui s'appuie non plus sur la connaissance de causes particulières, mais sur la multiplicité des causes possibles conduisant au même résultat. Ou alors, je peux m'appuyer sur le second principe de la

178 Pour un aperçu des différents types de prédictions réalisées (et confirmées) en biologie évolutive, voir Braude (1997).

thermodynamique qui affirme que les systèmes fermés évoluent toujours vers des états d'entropie croissante. En langage courant, cela veut dire que le désordre a tendance à augmenter. Dans ce cas, je n'ai pas besoin de connaître ni d'imaginer les causes pour en déduire la conséquence, mais je peux néanmoins être sûr que l'équilibre thermodynamique, c'est-à-dire l'état de désordre maximal, est à peu près inéluctable et que mon verre finira par se briser[179]. Je peux donc prédire une fin qui ne dépend pas d'une cause.

Considérons maintenant une goutte d'huile dans un verre l'eau — ou vice-versa. Non seulement je peux prédire qu'elle va former une bulle ou un rond, d'après mon expérience du phénomène, mais je peux ajouter qu'elle va le faire *pour* minimiser sa surface de contact avec le liquide environnant. Cette explication peut reposer soit sur l'intuition naïve qui dit que l'eau et l'huile se « repoussent », soit sur l'intuition physique qui dit que les systèmes tendent à évoluer vers leur état d'énergie minimale. Dans un cas comme dans l'autre, le raisonnement consiste à partir du résultat attendu et à prédire le moyen le plus efficient, parmi ceux disponibles, pour arriver à cette fin.

L'intuition naïve nous dit que si deux liquides se repoussent, ils vont avoir tendance à s'éloigner l'un de l'autre ; mais puisqu'ils sont en contact direct et ne peuvent pas se séparer physiquement, alors ils vont chercher à réduire au maximum ce contact ; l'eau va donc se ramasser sur elle-même et se mettre en boule, car la sphère est la forme qui présente la plus petite surface possible pour un volume donné.

L'intuition physique, quant à elle, nous dit qu'à l'interface entre deux liquides non miscibles se crée une tension ou énergie qui dépend de la surface et qui est minimale lorsque la surface est minimale ; or, puisque les systèmes tendent toujours vers les états d'énergie les plus faibles, on peut s'attendre à ce que la goutte adopte une forme sphérique pour réduire son énergie d'interface. Il s'agit bien ici d'une explication téléologique, car elle n'indique pas de mécanisme causal responsable de l'évolution du système et elle mentionne seulement son état final, en vertu duquel cette évolution est intelligible.

Cette seconde explication est sans doute approximative et incomplète, mais on ne peut pas dire qu'elle soit fausse, contrairement à la première, ni qu'elle implique une prédétermination de l'avenir et encore moins une intentionnalité naturelle ou surnaturelle, ni même qu'elle contrevienne aux lois connues de la physique. Bien au contraire, comme nous l'avons vu à la SECT. 4 et comme le reconnait Ernest Nagel (1961, p. 407), une grande partie de la physique classique et moderne peut être formulée à partir de principes comme celui de moindre action qui

179 Cela va dépendre en réalité du scénario cosmologique que l'on envisage pour la fin de l'univers.

affirment que les systèmes se comportent de manière à minimiser ou à maximiser certaines grandeurs.

Revenons-en à nos organismes capables de s'alimenter de plastique. La prédiction peut s'appuyer sur un principe intuitif général selon lequel les êtres vivants tendent à s'adapter à leur environnement physique ou évoluent de manière à maximiser cette adaptation. Le plastique est une source potentielle d'énergie chimique sous la forme de molécules organiques complexes et l'intuition nous dit que les êtres vivants tendent à exploiter, dans la mesure du possible, toutes les sources d'énergie à leur portée. Ces principes ne sont peut-être que des simplifications grossières de la théorie darwinienne et ils pourraient ne pas être toujours vérifiés ni vérifiables, mais ils nous permettent de formuler des hypothèses vraisemblables concernant le futur, le présent ou le passé, et d'envisager les moyens de leur réalisation. Ici, en partant de l'hypothèse que les êtres vivants seront un jour capables de dégrader le plastique pour en tirer de l'énergie, on peut supposer que cette évolution se fera par le biais d'une mutation génétique ou d'une exaptation, c'est-à-dire à travers l'apparition ou la réutilisation d'éléments (enzymes ou autres) n'ayant pas initialement été sélectionnés pour cela. On peut également supposer que cette évolution se produira chez des organismes comme les bactéries et les champignons pour au moins trois raisons : ils sont nombreux, ils peuvent évoluer de manière très rapide et ils sont déjà capables de dégrader des molécules similaires. De plus, on connaît l'antécédent des champignons capables de décomposer la lignine et la cellulose du bois et qui ont ainsi mis fin au carbonifère (Floudas et al., 2012). Ces hypothèses nous permettent notamment de mieux cibler la recherche des organismes appropriés, comme l'équipe qui a trouvé dans la forêt amazonienne un champignon capable de dégrader le polyuréthane en conditions anaérobies (Russell et al., 2011).

En biologie comme en physique, on peut donc réaliser des prédictions téléologiques de l'évolution d'un système à partir des principes généraux de ces disciplines et, en particulier, à partir des principes extémaux ou variationnels. Ces prédictions ne sont pas incompatibles avec une explication causale, mais elles ne dépendent pas d'une histoire causale particulière, contrairement aux explications mécaniques. Cela les rend à la fois plus générales et plus simples lorsque les histoires causales conduisant au résultat sont trop nombreuses ou trop compliquées.

3. Autres prédictions téléologiques en biologie

La téléologie permet de réaliser des prédictions biologiques qui seraient impraticables autrement. Si au lieu d'anticiper le résultat, on devait y arriver seulement par déduction à partir des phénomènes actuels, il faudrait alors par exemple envisager toutes les mutations génétiques

aléatoires possibles chez tous les organismes terrestres à un moment donné, voir si les gènes mutés codent pour des protéines ou sont susceptibles d'une quelconque autre activité biologique, voir le cas échéant si les protéines ou les activités en question contribuent ou pas à la *fitness* de l'organisme étant donné l'environnement dans lequel il vit, et répéter le processus autant de fois que nécessaire, de génération en génération, jusqu'à prédire la décomposition de la lignine, la photosynthèse et toutes les voies évolutives possibles de toutes les espèces sur une période de temps déterminée. La méthode est comparable à celle d'un joueur d'échecs qui voudrait calculer tous les déplacements possibles de toutes les pièces, tour après tour, jusqu'à arriver à trouver une combinaison victorieuse. Étant donné que chaque joueur a 20 possibilités de mouvement au début de la partie, le nombre de combinaisons s'élève à 400 pour le premier coup, à 20 000 pour le deuxième, et continue de s'incrémenter exponentiellement jusqu'à atteindre 10^{120} pour l'ensemble des parties possibles. Ce nombre est très largement supérieur à celui des atomes dans l'univers qui n'est « que » de 10^{80}, mais il est absolument infime par rapport aux possibilités combinatoires du vivant. Richard Dawkins (1998) rapporte par exemple que la probabilité de conformer une molécule d'hémoglobine au hasard à partir des vingt acides aminés les plus courants dans la nature, d'après le calcul de Isaac Asimov, s'élève à 10^{-190} ; et celle de former une cellule complète, d'après Michael Denton (1998), serait de 10^{-2000}. Si l'on considère en outre que certains progrès évolutifs impliquent la collaboration nécessaire de plusieurs organismes, comme l'endosymbiose à trois des premières cellules eucaryotes photosynthétiques (D. C. Price et al., 2012), on se rend vite compte de l'impossibilité d'un calcul « brut » ou « mécanique » pour formuler des prédictions biologiques. Bien entendu, en biologie comme aux échecs, on peut écrire des algorithmes qui réduisent drastiquement la quantité de calculs à effectuer en n'explorant par exemple que les branches de l'arbre combinatoire qui, après évaluation, semblent *a priori* les meilleures, c'est-à-dire les plus « prometteuses », et cette méthode permet aujourd'hui aux ordinateurs de battre les meilleurs joueurs d'échecs professionnels, mais elle est inabordable pour d'autres jeux comme le Go[180] et le jeu de la vie, le vrai.

Le raisonnement téléologique évite la plupart de ces calculs en commençant par la fin — ou par l'une des fins possibles — et en remontant vers la situation initiale. Aux échecs comme dans l'art militaire, on distingue la stratégie de la tactique. La première consiste à définir des objectifs et à dresser des plans sur le long terme à partir desquels on évalue ensuite les moyens (tactiques) à mettre en œuvre pour leur exécu-

180 Le programme AlphaGo, développé par Google DeepMind, a atteint en 2016 le niveau des meilleurs joueurs mondiaux, mais il utilise des réseaux de neurones profonds et des algorithmes d'auto-apprentissage.

tion. L'avantage évident de cette méthode est qu'elle permet de se concentrer sur un nombre limité de situations finales en ignorant les innombrables chemins qui n'y conduisent pas, et, parmi ceux qui restent, de n'étudier avec un certain détail que les plus probables et les efficients. Elle requiert une perception globale de la situation, de ses tendances, relations et lignes de force, et une grande capacité d'abstraction et de généralisation. En biologie comme dans d'autres domaines scientifiques, des méthodes similaires sont employées pour expliquer et pour prédire des phénomènes en tous genres. L'un des exemples les plus frappants est sans doute la prédiction et découverte ultérieure du comportement eusocial chez le rat-taupe nu.

En 1974, rapporte Stanton Braude (1997), le biologiste évolutionniste Richard D. Alexander pensait que les soins parentaux étaient un facteur déterminant dans l'évolution vers l'eusocialité chez les insectes, c'est-à-dire vers une organisation sociale comme celle des fourmis et des abeilles, avec un individu reproducteur et de nombreux travailleurs stériles. Or, si cette hypothèse était vraie, on pouvait s'attendre à ce que la même évolution se produise chez des vertébrés comme les oiseaux et les mammifères qui s'occupent aussi de leurs petits. Nonobstant, puisque les espèces de vertébrés sont beaucoup moins nombreuses et plus récentes que les insectes, on pouvait également s'attendre à ce qu'une telle évolution ne se soit pas encore produite. Alexander décida alors de faire comme si l'animal existait réellement. À partir de sa compréhension des forces sélectives impliquées dans l'évolution de l'eusocialité chez les insectes, il élabora un modèle en 12 points indiquant les caractéristiques que devrait avoir un vertébré eusocial s'il venait à exister. Il fallait par exemple que le nid fût difficile d'accès pour le protéger des prédateurs, susceptible d'être agrandi pour permettre l'accroissement de la population et proche d'une source de nourriture à la fois abondante et accessible sans risque. Cela conduisait à penser que les animaux devaient être complètement souterrains et qu'il s'agissait donc probablement de mammifères et plus particulièrement de rongeurs dont la principale source de nourriture seraient de larges racines et tubercules. En 1976, un autre chercheur suggéra que l'hypothétique rongeur ressemblait beaucoup au rat-taupe nu (*Heterocephalus glaber*) dont on ne connaissait pas jusque-là le comportement social. Alexander entra alors en contact avec une spécialiste de cet animal avec laquelle ils purent vérifier sur le terrain et en laboratoire que l'animal était effectivement eusocial et que les éléments de son modèle étaient corrects.

Le raisonnement qui le conduisit à cette découverte correspond à ce que nous disions plus haut : au lieu de chercher à prédire quelles espèces animales actuelles pourraient éventuellement évoluer vers l'eusocialité à partir de la situation actuelle, et au lieu de procéder à une étude exhaustive du comportement social des vertébrés pour vérifier si n'existait pas déjà, Alexander décida de commencer par la fin en s'imaginant le résultat

d'une telle évolution et en examinant les conditions ou les moyens nécessaires pour y arriver. Comme un général qui visualise la victoire et qui dresse ses plans de manière à la rendre possible et presque nécessaire, le biologiste élabora un modèle qui devait le conduire à l'animal recherché, fût-il réel ou hypothétique.

La prédiction d'Alexander n'est pas un fait isolé, puisque déjà Darwin avait prédit, entre autres choses, l'existence et les caractéristiques d'un insecte censé avoir coévolué avec une orchidée endémique de Madagascar (*Angraecum sesquipedale*) qui ne fut découvert qu'en 1903, soit vingt ans après la mort du savant, et dont la première observation en train de polliniser la plante n'eut lieu qu'en 2004.

Cela étant, le travail des biologistes ne consiste pas seulement à imaginer d'éventuels développements évolutifs ni à anticiper la découverte de nouvelles espèces. Bien souvent, leur travail consiste plutôt à décrire, à classer, à analyser et à manipuler les organismes ou les traits organiques qu'ils ont sous les yeux ; des activités qui contribuent à leur explication.

4. Catégorisation fonctionnelle et inférences

L'une des activités fondamentales de n'importe quelle science est la classification. De la chimie à la psychiatrie, en passant par l'astronomie, la géologie et la lingüistique, les scientifiques s'occupent de décrire et de classer leurs objets d'étude. Nous avons déjà évoqué au CHAP. V, SECT. 6.1, le rôle que joue la catégorisation dans la constitution des connaissances à travers les deux principes de Rosch : le *principe d'économie cognitive*, qui consiste à obtenir le maximum d'informations sur l'environnement avec un minimum d'effort cognitif, et le *principe d'exploitation de la structure du monde perçu*, qui s'appuie sur le fait que le monde se présente effectivement de manière structurée et pas sous forme d'attribus arbitraires ou imprédictibles. Ce deuxième principe permet de comprendre pourquoi la classification est d'une importance capitale pour les sciences : parce qu'elle rend possibles de nombreuses inférences.

En chimie, le tableau périodique de Mendeleïev a permis de prédire l'existence et les propriétés de plusieurs éléments inconnus et de corriger certaines erreurs concernant des éléments déjà connus à l'époque, car leur position dans le tableau correspond aussi bien à leurs propriétés physiques que chimiques, de sorte qu'un élément que son numéro atomique et sa configuration électronique placent dans la colonne des gaz nobles ou des métaux alcalins aura probablement des propriétés chimiques similaires à celles des autres éléments de sa catégorie.

En biologie, la classification darwinienne des êtres vivants repose sur la généalogie et la ressemblance, et il n'est pas étonnant que des espèces qui descendent d'ancêtres communs aient des traits qui se ressemblent,

de sorte que l'on aura tendance à trouver chez elles des caractères associés qui favorisent les inférences prédictives. Si nous savons qu'un individu appartient à un taxon (par exemple, celui des oiseaux) nous pouvons inférer un grand nombre de ses traits ou propriétés (ailes, plumes, bec, sang chaud, œufs, etc.). Comme dit Ernst Mayr (1997), les classifications « ont une grande valeur heuristique et explicative pour la plupart des branches de la biologie, comme la biochimie évolutive, l'immunologie, l'écologie, la génétique, l'éthologie, et la géologie historique ».

La taxinomie darwinienne ne s'appuie pas directement sur des catégories fonctionnelles. Cependant, dans la mesure où beaucoup de traits organiques sont effectivement porteurs de fonctions et dans la mesure où ces fonctions sont organisées pour contribuer ensemble au fonctionnement de l'organisme, il n'est pas étonnant qu'elles soient elles-mêmes associées de façon non arbitraire. Par exemple, si on sait qu'un individu a un cœur, que l'on peut comparer à une pompe, alors on peut prédire qu'il possède également des conduits ou vaisseaux (veines, artères), des échangeurs chimiques (poumons, branchies, peau) et des filtres (reins). De même, si l'on trouve une dent, on peut prédire que l'organisme en question devait être autotrophe et avoir un système digestif, car la fonction des dents est de saisir, de déchirer, de ronger ou de broyer des aliments. Il y a donc deux raisons principales pour lesquelles certains traits se trouvent souvent associés, la première est généalogique, la seconde est fonctionnelle.

Selon la définition que nous avons proposée au CHAP.VII, la fonction d'un trait est sa contribution à une fin du système auquel elle appartient. En général, d'autres traits ont aussi des fonctions qui contribuent à la même fin. Ces différents traits ou, plutôt, ces différentes fonctions ne sont pas indépendantes les unes des autres, mais organisées en systèmes plus ou moins complexes qui interagissent avec d'autres systèmes à plusieurs niveaux hiérarchiques au sein du tout intégré qu'est l'organisme. C'est leur intégration et leur organisation qui font que la finalité à laquelle chaque fonction contribue soit rendue possible. D'un côté, cela implique que chaque fonction est solidaire d'autres fonctions ; de l'autre, cela justifie le principe d'optimalité. Ce qui nous pousse à classer une structure dans une catégorie fonctionnelle, c'est notamment le fait qu'elle soit davantage appropriée pour cette fonction que pour une autre, et cela n'empêche pas une même structure d'appartenir à plusieurs catégories fonctionnelles, comme les plumes de l'oiseau qui lui permettent à la fois de voler, de s'isoler du froid, de se protéger de l'humidité et du soleil, de se camoufler, de communiquer, etc. Mais s'il existe une relation d'optimalité entre les fonctions et la finalité à laquelle ils contribuent, alors l'appartenance d'un trait à une catégorie fonctionnelle nous renseigne non seulement sur l'existence d'autres traits fonctionnels solidaires, mais nous apporte aussi des informations plus générales sur l'organisme porteur.

Par exemple, en examinant la forme des dents d'un animal, on peut déterminer son régime alimentaire. Celles du requin blanc, de l'éléphant et du rat ne sont manifestement pas adaptées aux mêmes types d'aliments, ce qui implique d'autres différences au niveau de la machoire, du crâne, du système digestif, etc. Souvenons-nous des lézards de l'île de Pod Mrčaru (voir p.125) qui, en devenant herbivores, ont subi des modifications morphologiques de la tête, de la machoire et de l'intestin.

Dans une étude de Kelemen et al. (2003) mentionnée plus haut, des enfants auxquels on montre des images de trois animaux différents sont non seulement capables de placer les animaux apparentés dans une même catégorie à partir de leur ressemblance globale, comme la loutre et le vison, mais aussi de tirer des informations sur leur comportement ou leur niche écologique à partir des traits fonctionnels qu'ils partagent avec des animaux non apparentés, comme les pattes palmées de la loutre et celles d'un oiseau aquatique.

En paléontologie et paléoanthropologie, l'interprétation fonctionnelle des restes fossiles complète leur étude descriptive et comparative pour déterminer leur place dans la classification phylogénétique, surtout lorsque d'autres données importantes comme la datation font défaut. La découverte des os d'un nouvel hominine, *Homo naledi*, présentait justement ce défaut (Berger et al., 2015), car les chercheurs ignoraient s'ils dataient de plus de 2 millions d'années ou de moins de 100 000 ans, ce qui leur posait un problème évident pour établir les liens de parenté avec *Homo sapiens* et d'autres espèces. La publication montre bien comment les auteurs s'appuyaient sur la signification fonctionnelle des structures anatomiques pour classer les fossiles découverts dans le genre Homo :

> « The overall morphology of *H. naledi* places it within the genus Homo rather than Australopithecus or other early hominin genera. The shared derived features that connect *H. naledi* with other members of Homo occupy most regions of the *H. naledi* skeleton and represent distinct functional systems, including locomotion, manipulation, and mastication. […] Locomotor, manipulatory, and masticatory systems have both historical and current importance in defining *Homo*, and *H. Naledi* fits within our genus in these respects. »

Et cette interprétation fonctionnelle leur permettait de formuler des prédictions concernant leur mode de vie, comme des phalanges très incurvées pour grimper aux arbres, et concernant leurs ancêtres, comme la bipédie et la manipulation d'objets[181] :

181 Les fossiles présentent un mélange de caractères modernes et primitifs : les pieds sont presque identiques à ceux des humains actuels, ce qui les rend aptes pour la marche bipède, les poignets, les pouces et les paumes des mains ont un aspect moderne, et rendent possible l'usage d'outils, mais les phalanges sont incurvées, comme pour grimper aux arbres, les épaules sont

> « This species combines a humanlike body size and stature with an australopith-sized brain; features of the shoulder and hand apparently well-suited for climbing with humanlike hand and wrist adaptations for manipulation; australopith-like hip mechanics with humanlike terrestrial adaptations of the foot and lower limb […]. We propose the testable hypothesis that the common ancestor of *H. Naledi*, *H. Erectus*, and *H. Sapiens* shared humanlike manipulatory capabilities and terrestrial bipedality […]. »

L'attribution de fonctions aux traits anatomiques, dans le cas de *Homo naledi*, semblent ne pas reposer sur une définition étiologique ni dispositionnelle. En effet, elle ne s'appuie pas sur l'étiologie pour déterminer la fonction, puisque l'histoire des ossements est inconnue, mais s'appuie à l'inverse sur leur morphologie et sur leur fonction présumée pour en tirer une hypothèse étiologique[182]. Elle ne résulte pas non plus d'une analyse fonctionnelle à la Cummins, laquelle partirait d'une capacité systémique arbitrairement choisie par l'observateur pour déterminer la contribution causale de ses parties, mais s'appuie plutôt sur la morphologie des os pour en inférer les capacités globales auxquelles ils devaient contribuer. La méthode des auteurs semble dès lors se rapprocher davantage de la rétro-ingéniérie d'un système dysfonctionnel, comme le mécanisme d'Anticythère, où l'on examine d'abord les parties et leurs relations pour chercher les fins dont elles pourraient être les moyens, suivant le principe d'optimalité, et en inférer ensuite l'origine causale la plus probable. De cette manière, on fait coïncider la fonction systémique avec la fonction étiologique.

De façon générale, les fonctions contribuent à la classification en biologie en ce qu'elles permettent certaines inférences et généralisations qu'il serait difficile — si ce n'est impossible — de réaliser autrement. En premier lieu, l'attribution fonctionnelle est le meilleur moyen dont nous disposions pour classer dans une même catégorie des traits qui sont présents chez différentes espèces, comme les cœurs, et dont nous reconnaissons intuitivement l'unité (fonctionnelle) malgré leur diversité (morphologique) :

---

simiesques, les molaires sont modernes, les racines des prémolaires sont primitives, etc.

182 L'âge présumé de *Homo naledi* est d'environ 2 millions d'années. Il appartiendrait au même groupe que *Homo erectus*, Néandertal et les humains modernes, et il serait plus proche de *Homo erectus* que *Homo habilis*, c'est-à-dire qu'il pourrait faire partie de nos ancêtres. Cependant, tous les spécialistes ne sont pas d'accord sur ce point.

> « For instance, "heart" cannot be defined except by reference to the function of hearts because no description purely in terms of morphological criteria could demarcate hearts from non-hearts. Biologists need a category that ranges over different species, and hearts are morphologically diverse […]. » (Neander, 1991a, p. 180)

Dans la plupart des cas, les traits fonctionnels sont aussi des homologues, comme les cœurs, puisqu'ils procèdent d'un ancêtre commun. L'attribution fonctionnelle permet alors d'ajouter un niveau de précision supplémentaire dans la classification, car tous les traits homologues n'ont pas la même fonction et certains d'entre eux sont vestigiaux. Neander et Rosenberg revendiquent ainsi l'existence et l'importance des catégories homologues-fonctionnelles en biologie[183] (Neander, 2002, 2010; Rosenberg & Neander, 2009).

Dans d'autres cas, en particulier lorsqu'il y a convergence évolutive, les traits fonctionnels sont seulement analogues, comme les yeux, les écailles, les nageoires, les ailes et les queues préhensiles. On peut alors attribuer une même fonction et former une même catégorie biologique à partir de structures qui n'ont pas la même origine ni la même histoire, mais qui constituent des réponses évolutives à des problèmes écologiques similaires. D'après Paul Griffiths, les classifications fonctionnelles complètent les classifications cladistiques à un degré de généralité supérieur qui permet de mieux saisir les dynamiques de l'évolution, ce qui fait leur valeur :

> « If functional classifications are to be of value in biology it must be because of their superior generality—the fact that they unite disjunctions of cladistic homologues. […] So long as some generalizations are made about functionally classified items, introducing a functional classification will add to our understanding of the dynamics of evolution. » (Griffiths, 1994, p. 213-7)

À ce niveau de généralité, on peut considérer que les catégories fonctionnelles ont une valeur explicative non causale, car elles permettent de rendre compte de phénomènes causalement et historiquement distincts qui aboutissent au développement de moyens similaires pour répondre à des problèmes similaires. Si les oiseaux, les chauves-souris et les papillons ont des ailes pour voler, c'est parce que la sustentation dans l'air est un problème physique universel qui peut en partie être résolu grâce à une surface portante suffisamment grande et légère par rapport à la taille et au poids de l'animal, et à des battements cadencés. Si les requins et les dauphins ont des morphologies similaires, c'est parce que les déplacements dans l'eau posent un problème physique que les formes hydrodynamiques contribuent à résoudre. Les traits fonctionnels analogues sont donc le produit de la rencontre de deux ensembles de

183 Il existe une polémique sur ce point, voir CHAP. I, note 13, p. 64.

contraintes, celles issues de la morphologie d'origine et celles de la niche écologique. Au niveau de chaque espèce, l'explication du trait est duale : généalogique et fonctionnelle. Au niveau supérieur, elle est purement fonctionnelle. Dans un cas comme dans l'autre, l'explication peut être formulée en termes d'optimalité : les ailes et les nageoires comptent sans aucun doute parmi les meilleures moyens disponibles, étant donné les contraintes de départ, pour se déplacer en l'air et dans l'eau.

De nombreux philosophes et scientifiques refuseraient probablement de considérer cela comme une explication à défaut de connaître les causes ou les mécanismes qui expliquent comment les dauphins ont acquis leurs nageoires. Mais il existe plusieurs conceptions non causales de l'explication. Parmi elles, la conception de l'unification théorique formulée par Kitcher (1981, 1985) pourrait être appropriée pour ce genre d'explication qui, malgré sa simplicité, permet de rendre compte d'un très grand nombre de phénomènes différents et d'en tirer un très grand nombre d'inférences. Mais nous allons voir que ce n'est pas la seule forme possible ni la seule fonction envisageable des explications téléologiques en biologie.

5. Acceptabilité et généralité des explications téléologiques

Du point de vue de la psychologie cognitive, les explications ont plusieurs fonctions dont la principale est d'aider à prédire et à contrôler les événements futurs et, plus généralement, à guider le raisonnement inférentiel, en particulier dans la généralisation des connaissances acquises à des cas nouveaux (Keil, 2006; Lombrozo, 2006). Nous avons vu dans les sections précédentes comment le raisonnement téléologique permet de réaliser des prédictions et des inférences inductives. Nous voulons maintenant comprendre ce qu'apporte la téléologie à l'explication en biologie et ce qui fait que les explications téléologiques y soient jugées valables.

Une étude menée par Tania Lombrozo et Susan Carey (2006) sur des étudiants universitaires de Harvard a montré que l'une des conditions d'acceptabilité des explications téléologiques est la généralité, c'est-à-dire que pour être acceptable une explication téléologique doit pouvoir être généralisée à des situations nouvelles :

> « We found that adults accept teleological explanations when two conditions obtain : (a) the function invoked in the explanation played a causal role in bringing about what is being explained and (b) the process by which the function played a causal role seems general, in the sense that it conforms to a predictable pattern. »

Dans l'une des expériences de cette étude, les sujets étaient divisés en deux groupes. Ceux du premier groupe recevaient une feuille avec une histoire causale narrant la création d'un trait et devaient évaluer l'acceptabilité de différentes explications (mécanique, téléologique et intentionnelle) de l'existence de ce trait. Ceux du deuxième groupe recevaient la même histoire causale suivie des mêmes explications à évaluer, mais précédée de deux autres histoires causales similaires. Les résultats indiquent que les sujets étaient beaucoup plus enclins à accepter la validité de l'explication téléologique lorsqu'ils voyaient que celle-ci était généralisable à des situations similaires. À partir de ces résultats, les auteurs formulaient l'hypothèse que l'une des fonctions psychologiques de l'explication est d'identifier les facteurs « exportables », c'est-à-dire généralisables et réutilisables, notamment pour formuler des prédictions.

Cette hypothèse a été reprise et confortée par une étude postérieure de Tania Lombrozo (2010) sur des étudiants de Berkeley visant à montrer que l'identification des causes d'un événement ou d'un trait est liée au type d'explication qu'on en donne (mécanique ou téléologique) : selon l'explication que l'on donne d'un même fait, les causes jugées responsables de ce fait ne seront pas les mêmes.[184] Précédemment, la même auteure avait montré que le type d'explication affecte la catégorisation et les inférences qui en dérivent (Lombrozo, 2009). L'une des conclusions qu'elle tire de ces études est que les explications mécaniques et téléologiques sont complémentaires en ce qu'elles permettent des généralisations et des prédictions différentes qui capturent des aspects différents de la réalité. De plus, le choix du type d'explication dépend de la chose à expliquer, de sorte que la téléologie est plus appropriée pour les systèmes qui manifestent une équifinalité (ce qui est justement le cas des systèmes biologiques) :

> « Different modes of explanation are useful because they capture different but real regularities in the environment. Reasoning teleologically – in terms of functions and design – makes it possible to capture relationships between behavior and outcomes that would be very difficult to express in terms of purely physical variables. Likewise, reasoning mechanistically – in terms of causal processes – makes it possible to capture relationships that would be very difficult to express in terms of intentional or functional variables. The fact that each mode of explanation employs variables that support

184 Les explications mécaniques requièrent que deux critères soient satisfaits pour identifier une relation causale : la dépendance et le transfert. Le premier signifie que l'effet dépend de la cause. Le second signifie qu'il existe une relation « mécanique » directe entre l'un et l'autre, comme le transfert de la quantité de mouvement d'une boule de billard à une autre lors d'un choc. Les explications téléologiques identifient les causes selon le critère de dépendance mais pas celui de transfert.

> particular generalizations suggests that variable selection is itself subject to a criterion of exportability. It is therefore unsurprising that explanatory mode should have consequences for the judged exportability of particular relationships, and hence for causal ascriptions. Moreover, one should expect reasoners to spontaneously adopt the mode of explanation that supports greater exportability, such as a teleological mode when assessing an equifinal system. » (Lombrozo, 2010, p. 329)

Comme l'auteure s'appuie sur la conception étiologique des explications téléologiques défendue par Larry Wright (voir Lombrozo & Carey, 2006), les études citées ont en commun le fait de présenter aux sujets des histoires causales et de formuler des questions portant sur la présence d'un trait — des questions pour lesquelles différentes explications sont alors proposées. Les explications téléologiques y sont dès lors conçues comme des explications causales qui se distinguent des explications mécaniques en ce qu'elles citent les effets de la présence du trait et pas les mécanismes causaux qui en sont responsables.

La généralité ou l'« exportabilité » de l'explication consiste ici dans le fait que dans des histoires causales portant sur des objets différents, on retrouve un même schéma où les effets d'un trait sont rétroactivement responsables de la présence ou des propriétés de ce trait. On explique ainsi de la même manière pourquoi tel animal a de grandes oreilles, tel autre a de grands yeux ou encore pourquoi tel artefact brille dans le noir.

Par ailleurs, Lombrozo et Carey justifient leur hypothèse sur la fonction et les conditions d'acceptabilité des explications téléologiques, entendues en termes étiologiques, en faisant référence aux conceptions de l'explication scientifique développées par Michael Strevens (2004), James Woodward (2003) et John Campbell (2008). La psychologie cognitive semble donc reconnaître que les explications téléologiques sont des explications authentiques, et cette justification paraît compatible avec certaines des plus récentes théories philosophiques de l'explication[185].

185 Conformément à la conception de Strevens, les explications téléologiques permettraient d'identifier les causes pertinentes de la chose à expliquer, c'est-à-dire celles qui « font la différence » (*difference-makers*), et elles invoqueraient des modèles causaux très généraux, c'est-à-dire applicables à un très grand nombre de situations et de systèmes physiques différents. Conformément à la conception de Woodward, les explications téléologiques permettraient d'identifier des relations de dépendance invariante face à des variations contrefactuelles, car les systèmes et les processus téléologiques se caractérisent justement par l'équifinalité, c'est-à-dire par le fait de pouvoir atteindre une même fin par des moyens différents. Conformément à la conception de Campbell, les explications téléologiques permettraient d'identifier des relations de dépendance causale assez particulières, notamment au sens où la cause varie en fonction du résultat et où des variations paramétriques de la cause induisent des variations correspondantes de l'effet.

Cependant, la justification en question s'appuie sur une conception causale de l'explication et de la téléologie. Or, nous avons défendu jusqu'ici une conception non causale de cette dernière. Nous devons donc en proposer une justification différente.

C'est ce que nous allons faire dans la prochaine section à partir de la notion d'invariance. Celle-ci joue un rôle central dans la théorie de l'explication de James Woodward et en particulier dans son application à la biologie, car elle lui permet de justifier de caractère explicatif de généralisations biologiques comme la loi de Mendel. Ce que nous allons montrer, c'est que la justification de cet auteur est extensible aux explications téléologiques dans la mesure où elles permettent d'identifier des invariants différents de ceux des explications mécaniques et historiques.

6. Valeur explicative, interventions et invariance

D'après Woodward (1997, 2000, 2003), une généralisation qui décrit une relation entre plusieurs variables est explicative si elle est invariante, et elle est invariante si elle continue de s'appliquer — c'est-à-dire si elle demeure stable ou inchangée — lorsqu'elle est soumise à un certain nombre de variations ou d'interventions. Par exemple, la loi de la gravitation de Newton implique une relation entre les masses, la distance et la force qui demeure inchangée aussi bien face à des changements dans les variables qui n'interviennent pas dans l'équation, comme la position, la vitesse ou la charge électrique, que face à une large gamme de variations de leurs masses et de leur distance. Ce second type de modification, qui affecte directement les variables de l'équation, est ce que cet auteur appelle une intervention, laquelle n'implique pas nécessairement une action humaine.

Si une généralisation invariante est explicative, c'est parce qu'elle dévoile l'existence d'une relation de dépendance entre l'*explanandum* et l'*explanans* telle que si les facteurs cités dans l'*explanans* étaient ou avaient été différents, l'*explanandum* serait lui-aussi différent :

> « On my view, to explain an explanandum is to show how changes in it countefactually depend on changes in the factors cited in the explanans, or to express the same idea in a slightly different way, to show how the explanandum would have been different if the factors cited in its explanans had been different in various ways. » (J. Woodward, 2001, p. 5)

Dans ce contexte, Woodward défend une conception causale de l'explication où la notion de cause est entendue au sens large comme une dépendance contrefactuelle. Ce qui la distingue d'autres théories contrefactuelles de la causalité, c'est le fait que la relation de dépendance causale

doive être invariante par rapport à des interventions. Par exemple, on peut établir une relation de dépendance contrefactuelle entre les mesures d'un baromètre, la pression atmosphérique et la météo, de sorte que si la mesure du baromètre augmente rapidement, alors il fera beau, tandis que si elle baisse rapidement, un orage va se produire. Pourtant, ce n'est pas l'aiguille de l'appareil qui fait la pluie et le beau temps et on ne peut pas recourir à lui pour expliquer la survenue de l'orage. En effet, la relation n'est pas invariante par rapport à des interventions sur le baromètre, car on ne peut pas contrôler ni manipuler la météo en manipulant celui-ci. En revanche, on peut expliquer les mouvements de l'aiguille à partir des variations de la pression atmosphérique, car il existe entre les deux une relation invariante telle que l'on peut contrôler les mesures en intervenant sur la pression. C'est donc la météo qui explique les mouvements de l'aiguille du baromètre et pas l'inverse.

Les lois de la physique sont invariantes en ce sens, mais une généralisation peut être invariante sans être une loi. C'est ce qui permet à cette théorie d'être applicable aussi bien à la physique qu'aux sciences sociales et à la biologie où certains affirment qu'il n'y a pas de lois véritables. On reproche en effet aux généralisations biologiques d'être contingentes et truffées d'exceptions. D'après Woodward (2001), la question n'est pas de savoir si ces généralisations sont des lois, mais si elles sont explicatives, c'est-à-dire si elles sont invariantes de la façon indiquée ci-dessus. Or, en analysant depuis cette perspective la loi de Mendel et celle de Hardy-Weinerg, il montre comment et par rapport à quelles interventions contrefactuelles ces généralisations sont effectivement invariantes et, par conséquent, explicatives.

En ce qui nous concerne, nous voulons montrer que cette conception de l'explication est applicable à la téléologie biologique tout en défendant que les explications téléologiques ne sont pas causales. Dans l'expérience de Gergely et Csibra, les sujets expliquent le comportement du petit cercle à partir du principe d'action efficiente qui lie le comportement, la finalité et les contraintes de la situation. Le petit cercle se dirige en ligne droite ou saute par dessus le rectangle pour rejoindre le grand cercle (parce que c'est le meilleur moyen d'atteindre cette fin dans ces circonstances). Pour que cette explication téléologique soit valable, du point de vue de la théorie de Woodward, il faut qu'elle soit invariante sous des interventions contrefactuelles, de sorte que l'on puisse savoir ce qui se serait passé si les choses avaient été différentes. Or, c'est précisément ce que l'expérience met en évidence : lorsque l'on change la situation, en ajoutant ou en retirant un obstacle, les sujets s'attendent à un changement du comportement du petit cercle tel que la relation entre les moyens et la fin ne soit pas changée, c'est-à-dire tel que l'efficience soit préservée.

L'explication téléologique maintient aussi la même relation de dépendance asymétrique de l'*explanandum* envers l'*explanans* que l'on observe dans l'explication causale. Prenons l'exemple de deux boules de billard. La première se dirige vers la seconde, la frappe et celle-ci se met en mouvement. Du point de vue de l'explication causale, le mouvement de la première boule et le choc élastique (les causes) font partie de l'*explanans*, tandis que le mouvement de la seconde (la conséquence) est l'*explanandum*. Si les causes avaient été différentes, c'est-à-dire si la première boule ne s'était pas dirigée vers la seconde ou si elle ne l'avait pas frappée, alors les conséquences auraient été différentes, car elle n'aurait pas bougé. En revanche, si les conséquences avaient été différentes, la deuxième boule n'ayant pas bougé — parce qu'elle était fixée à la table, par exemple —, cela n'aurait rien changé aux causes, c'est-à-dire au fait que la première s'est déplacée vers elle et l'a frappée. C'est l'*explanandum* qui dépend de l'*explanans* et pas l'inverse.

Du point de vue de l'explication téléologique, le déplacement de la deuxième boule (la fin) est l'*explanans* et celui de la première (le moyen) l'*explanandum*, car celle-ci se déplace vers celle-là pour la mettre en mouvement. Si la fin avait été différente, c'est-à-dire par exemple si ce n'était pas la deuxième boule mais une autre qui devait être mise en mouvement, alors le moyen aurait été différent, car la première n'aurait pas été dirigée vers la seconde et ne l'aurait pas frappée. En revanche, si les moyens avaient été différents, cela n'aurait pas changé la fin. C'est donc ici aussi l'*explanandum* qui dépend de l'*explanans* et pas l'inverse.

Cette relation asymétrique remplit une condition additionnelle de la théorie de Woodward qui est que l'on puisse contrôler l'*explanandum* en manipulant l'*explanans*. Du point de vue de la relation causale, on peut en effet contrôler les conséquences en jouant sur les causes. Du point de vue de la relation instrumentale, on peut de même contrôler les moyens en modifiant la fin et les contraintes de la situation. En effet, si on demandait à un joueur de billard de mettre la deuxième boule dans le trou le plus proche, il faudrait qu'il frappe la première avec une force et une direction bien particulières. Si on lui demandait de la mettre plutôt dans un autre trou, il faudrait qu'il la frappe avec une force et une direction différentes. Mais le choix des moyens dépend aussi des autres boules présentes sur la table. En les changeant de place, on peut par exemple forcer le joueur à utiliser les bandes au lieu de frapper en ligne droite.[186] En manipulant la fin et les contraintes de la situation (*explanans*), on contrôle donc effectivement les moyens (*explanandum*).

186 Bien que la finalité dépende ici des intentions d'un agent, on pourrait facilement remplacer cet exemple par n'importe quel autre n'impliquant pas d'agent intentionnel, comme le petit cercle de l'expérience de Gergely et Csibra.

On pourrait croire que les explications téléologiques diffèrent des explications causales en ce qu'elles sont seulement qualitatives, comme quand on dit que les solides tombent pour rejoindre leur lieu naturel, mais cette appréciation est inexacte, car elles peuvent aussi avoir une valeur quantitative. Nous avons mentionné plus haut (voir SECT. 4, p. 312) l'exemple de certains corbeaux du Canada qui laissent tomber des bulots sur les rochers depuis une certaine hauteur pour en briser la coquille et en manger la chair. Or, il est possible de donner à cette explication téléologique une précision numérique en calculant la hauteur optimale pour briser la coquille avec le moindre effort étant donné les contraintes physiques de la situation. Cette hauteur est d'environ 5 mètres et correspond précisément au comportement des corbeaux. Si l'augmentation des températures moyennes ou l'acidification des océans modifiaient la croissance des mollusques et la formation des coquilles, la hauteur optimale depuis laquelle lâcher les bulots serait modifiée et on pourrait prédire une évolution correspondante du comportement des oiseaux. La relation est donc invariante sous des interventions contrefactuelles. De façon générale, dans la mesure où les explications téléologiques sont guidées par le principe d'action efficiente, c'est-à-dire par un principe d'optimalité, elles peuvent en principe recevoir une expression numérique.

De même que la loi de Newton établit une relation invariante entre la force gravitationnelle, les masses et la distance, le principe d'action efficiente établit une relation invariante entre la fin, les moyens et la situation. Cependant, on pourrait croire que les lois physiques ont une portée universelle tandis que les explications téléologiques sont seulement particulières ou spécifiques, celle que nous venons de voir n'étant par exemple valable que pour les corbeaux du nord-ouest du Canada. C'est faux. Non seulement l'explication précédence est applicable à d'autres organismes, comme le gypaète barbu, un vautour qui brise de grands os en les laissant tomber sur les rochers pour se nourrir de leur moelle, ou l'aigle royal, qui fait de même avec les tortues pour briser leur carapace, mais elle appartient à un type d'explication plus général qui est lié à l'efficience du comportement alimentaire, c'est-à-dire essentiellement au rapport entre la valeur nutritive d'un aliment et l'effort nécessaire pour l'obtenir, et qui s'applique aussi bien aux fourmis (Sumpter & Beekman, 2003) qu'aux humains (Hill, Kaplan, Hawkes, & Hurtado, 1987), car il semble évident que tout organisme vivant doit tirer de sa nourriture un bénéfice supérieur à ce qu'elle lui coûte. Le détail de la balance coûts-bénéfices va dépendre de chaque situation particulière, avec des stratégies extrêmement variées d'un organisme à un autre, certaines étant certainement plus efficientes que d'autres et aucune n'étant optimale en termes absolus, mais le principe explicatif qui rend compte de la diversité infinie des situations particulières est, quant à lui, universel.

Si l'on considère la définition minimale du concept de relation causale que formule Woodward (2010), on constate que la relation moyens-fins en satisfait les conditions :

> (**M**) *X* causes *Y* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *X* (and no other variable) were to occur in *B*, then *Y* or the probability distribution of *Y* would change.

En effet, si *X* représente la fin, *Y* les moyens et *B* la situation, alors nous avons vu plus haut qu'une intervention modifiant *X* (et aucune autre variable) dans *B* modifierait *Y* ou la distribution de probabilité de *Y*.

Les fins devraient-elles pour autant être considérées comme des causes, suivant cette définition ? Non, car il y a plusieurs raisons qui nous retiennent de le faire. L'une d'elles tient au type de relation entre les termes. Il nous semble en effet que la relation causale est une relation univoque entre la cause et son effet, au sens où les mêmes causes dans le même contexte produisent toujours les mêmes effets. En revanche, les relations téléologiques sont multivoques, au sens où une même fin est compatible avec (nous ne dirons pas qu'elle « provoque ») des moyens différents. Certes, tous les moyens ne se valent pas, certains étant manifestement plus simples ou plus faciles, ou plus efficients que d'autres, mais rien n'empêche que plusieurs d'entre eux soient équivalents selon l'un ou l'autre de ces critères, ce qui implique que les moyens sont sous-déterminés par les fins.

Le domaine biologique illustre à merveille cette sous-détermination, car pour atteindre une même fin, comme l'alimentation ou plus généralement le métabolisme, la diversité des moyens employés par les êtres vivants est virtuellement infinie ; et bien que les circonstances ne soient jamais exactement les mêmes, cette diversité est aussi en grande partie le fruit de contingences historiques. Or, l'une des restrictions qu'ajoute l'auteur à sa définition minimale est la reproductibilité :

> [T]he account of claims of the form *X causes Y* that I propose is restricted to cases in which the relationship between *X* and *Y* is general or reproducible in the sense that *Y* exhibits some sort of systematic response when the same changes in the value of *X* are repeated, at least in the right circumstances. […] Without reproducibility, the counterfactual claims on which my account relies will not be obviously true. (J. Woodward, 2003, p. 42)

Dans le domaine biologique, la reproductibilité est un problème, car rien ne nous garantit que l'histoire de la vie se répéterait de la même façon si elle venait à se reproduire. Si l'on considère une relation causale proximale entre deux variables dans un contexte relativement simple, comme l'exemple des boules de billard, il est facile de voir que les mêmes

causes produisent systématiquement les mêmes effets. Si l'on considère l'organisation et l'évolution du vivant, c'est beaucoup moins évident. Nonobstant, on peut penser que le problème réside ici dans la complexité ou la complication des relations causales et dans l'intervention à tous les niveaux de phénomènes aléatoires, de sorte que le principe n'est pas remis en question. Si l'on considère maintenant la relation entre les fins et les moyens, on voit que dans certains contextes simples, comme l'expérience de Gergely et Csibra, les mêmes fins conduisent aux mêmes moyens. Mais, même dans un contexte simple, une même fin peut parfois être réalisée par deux moyens également efficients. On peut alors expliquer l'existence d'un trait biologique ou d'un comportement donnés avec une certaine fonction dans la mesure où celle-ci est le moyen d'une fin comme l'alimentation ou la respiration, mais pour expliquer que ce soit précisément ce moyen là et pas un autre qui réponde à cette fin, il faut recourir à l'histoire et au hasard. En d'autres termes, on peut expliquer que l'organisme *O* soit doté du trait fonctionnel *Y* (des poumons) pour réaliser la fin *X* (l'oxygénation du sang), car si *X* était absent *Y* serait absent et si *X* était différent *Y* serait différent (c'est le principe de l'adaptation biologique), mais la même explication s'applique à l'organisme *O'* doté du trait *Y'* (des branchies) et à l'organisme *O"* doté de *Y"*, et il n'est pas sûr que les mêmes variations de *X* impliqueraient toujours les mêmes variations de *Y*, *Y'* et *Y"*. Ce qui est sûr, en revanche, c'est que si un organisme quelconque a besoin d'oxygène, alors il doit disposer d'un moyen efficient pour l'acquérir. La fin (*X*) explique le moyen (*Y*), quelle que soit la forme particulière qu'adopte ce dernier et quelle que soit l'histoire causale l'ayant conduit à l'adopter. Comme nous le disions à la section 1 (p. 305), les explications téléologiques font abstraction des causes et sont en ce sens plus générales. On pourrait dire aussi qu'elles se situent à un niveau d'explication plus élevé.

D'un autre côté, nous avons des raisons de penser qu'il existe néanmoins une certaine forme de relation systématique, fût-elle seulement probabiliste, entre les moyens et les fins. La principale de ces raisons est la convergence évolutive. Si des organismes très différents par leur histoire et leur morphologie finissent par trouver des solutions analogues pour les mêmes problèmes ou des problèmes similaires, ce n'est peut-être pas seulement l'effet du hasard. C'est peut-être aussi parce que ces solutions-là comptent parmi les meilleures possibles pour y répondre. Les lois de l'hydrodynamique sont universelles et certaines formes sont objectivement meilleures que d'autres pour se déplacer dans l'eau le plus rapidement possible, que ce soit pour un poisson, un mammifère, un oiseau ou un sous-marin nucléaire. C'est la raison pour laquelle on peut expliquer fonctionnellement un trait biologique indépendamment des contingences de son histoire évolutive et indépendamment même de la théorie darwinienne. Et c'est l'une des raisons pour lesquelles les explications téléofonctionnelles sont indispensables à la biologie, parce qu'elles

permettent d'identifier des invariants différents de ceux qu'identifient les explications causales.

7. Valeur explicative des attributions fonctionnelles

Selon la définition que nous avons proposée au CHAP. VII, la fonction d'un trait est une conséquence de la présence ou de l'activité de ce trait qui contribue à une fin du système auquel elle appartient. Selon le principe d'action rationnelle ou efficiente, les systèmes téléologiques sont censés employer pour atteindre leurs fins les moyens les plus efficients qui leur soient disponibles étant donné les contraintes de la situation. Ces contraintes incluent les autres éléments du système et son organisation interne, ainsi que les conditions et les circonstances qui lui sont externes.

Attribuer une fonction, c'est donc donner une explication causale partielle d'une propriété ou d'une capacité d'un système (la fin), à partir de la présence ou de l'action de l'un de ses éléments (la fonction du trait) en interaction avec d'autres éléments (les fonctions d'autres traits) dans un contexte donné (l'organisation du système et son environnement). Par exemple, on explique la capacité de l'organisme à alimenter les cellules en dioxygène et en nutriments à partir de l'action de pompage du sang que réalise le cœur dans le système circulatoire, en interaction avec les vaisseaux, les poumons, etc., dans un environnement où du dioxygène atmosphérique et des aliments sont accessibles.

Attribuer une fonction, c'est aussi donner une explication téléologique de la présence ou de l'action d'un élément (le trait) dans un système donné, à partir de la finalité de ce système et du principe d'action efficiente. Si l'on admet qu'un organisme vivant peut avoir des fins, alors la présence et l'activité d'un trait organique se justifie dans la mesure où elle constitue un moyen efficient de ces fins. Par exemple, on explique que les vertébrés aient un cœur parce qu'ils doivent apporter aux cellules les dioxygène et les nutriments dont elles ont besoin, parce que le système circulatoire est une très bonne solution pour résoudre ce problème (et c'est celle qu'ils emploient effectivement), parce que le système circulatoire requiert un mécanisme de pompage et, finalement, parce que le cœur est une très bonne pompe dans ce système.

Les explications causale et téléologique impliquent l'une et l'autre la décomposition d'un tout en ses parties, mais pas de la même façon. Dans le premier cas, on explique une capacité du tout (l'alimentation des cellules) à partir de l'organisation de ses parties (le système circulatoire) et du rôle (de pompe) que joue l'une d'elles (le cœur) dans cette organisation. Dans le deuxième cas, on explique la présence d'une partie (le cœur) et son activité (de pompage) à partir de la finalité du tout (la circu-

lation du sang pour alimenter les cellules) et de son organisation (le système circulatoire).

Par ailleurs, bien que les explications précédentes impliquent l'une et l'autre des relations causales entre les parties et le tout, elles ne leur imposent pas les mêmes exigences. Dans le premier cas, il suffit qu'une partie soit causalement *efficace* pour qu'elle contribue à expliquer les capacités du tout. Dans le second, il faut aussi qu'elle soit causalement *efficiente*. Ainsi, pour expliquer causalement l'alimentation des cellules du corps, il suffit que le cœur soit effectivement capable de pomper le sang dans le système circulatoire, quelle que soit la quantité de ressources qu'il utilise pour arriver à ce résultat. En revanche, pour expliquer téléologiquement la présence et l'activité du cœur dans l'organisme, il faut non seulement qu'il soit capable de pomper le sang, mais aussi qu'il le fasse de manière efficiente, c'est-à-dire qu'il consomme moins de dioxygène que ce qu'il apporte au reste de l'organisme. Plus le rapport entre les coûts et les bénéfices est faible, plus la relation causale entre les moyens et les fins est efficiente. Chez l'humain, ce rapport est de 1/10 pour le dioxygène, l'activité cardiaque représentant à peu près 11 % de la consommation totale de l'organisme au repos. Si l'on arrivait à greffer un cœur d'éléphant dans un corps humain et que ce dernier y survive, sa consommation au repos serait si grande que le seul fait d'avoir un cœur battant serait épuisant pour l'organisme. Il serait efficace et pourrait même se voir attribuer une fonction au sens de Cummins, mais on ne pourrait pas expliquer sa présence en termes téléologiques.

Pour que les explications précédentes soient valables selon la théorie de Woodward, il faut qu'elles révèlent une relation de dépendance contrefactuelle invariante entre l'*explanandum* et l'*explanans*. Dans le cas de l'explication causale, il est facile de montrer que l'apport de nutriments et de dioxygène aux cellules dépend de l'activité du cœur, car un arrêt cardiaque prolongé conduit presque nécessairement à une anoxie, et parce que l'on peut contrôler partiellement l'oxygénation des tissus en intervenant sur la fréquence cardiaque.

Dans le cas de l'explication téléologique, il est également facile de montrer que l'activité cardiaque dépend de l'apport de nutriments et de dioxygène aux cellules, car la fréquence des battements varie avec la demande des tissus, de sorte que l'on peut contrôler partiellement la première en intervenant sur la seconde, par exemple en augmentant l'effort musculaire, la masse de l'animal, la thermogenèse, etc. On peut aussi intervenir sur d'autres variables, comme la pression sanguine ou la quantité de dioxygène atmosphérique. Lorsque l'altération de la demande se prolonge dans le temps, chez les sportifs de haut niveau et chez les personnes souffrant d'hypertension chronique, d'autres modifications peuvent se produire, comme une hypertrophie du ventricule gauche qui augmente, à court et moyen terme, la capacité cardiaque. On constate aussi que chez les individus d'une même espèce, la taille du cœur est en

relation avec la taille de l'animal, et que, chez des espèces différentes, ce ratio dépend aussi en partie de leur mode de vie, c'est-à-dire notamment de leur activité physique. Dans la plupart des cas, ces variations de la capacité cardiaque, qui passent par des variations de sa taille et de sa fréquence, correspondent à des variations de la demande de l'organisme et aboutissent à un même résultat, à savoir : à l'équilibre entre l'apport et la demande. Cet équilibre est un invariant.

La modification de la capacité cardiaque n'est pas la seule manière d'arriver au même résultat. L'organisme est en effet capable de maintenir l'équilibre entre l'apport et la demande sans modifier l'activité cardiaque ou de façon complémentaire à cette modification, par exemple en augmentant le nombre de globules rouges, en faisant varier la pression artérielle, la fréquence respiratoire, le métabolisme basal, etc. Si à l'équifinalité de ces différentes mesures on ajoute le fait qu'elles sont généralement très efficientes, il ne devrait faire aucun doute que ce sont là des moyens d'une même fin.

Conformément à la définition téléologique que nous avons proposée, les battements du cœur ont pour fonction de faire circuler le sang dans le système circulatoire parce que :

(1) la circulation est une conséquence de l'activité cardiaque,

(2) elle contribue (de manière efficiente) à une fin du système qui est de répondre aux demandes énergétiques de l'organisme,

(3) elle constitue un élément essentiel du système, c'est-à-dire que sa contribution n'est pas accidentelle.

À titre de comparaison, on peut expliquer causalement la capacité de l'organisme à produire des sons à partir du bruit des battements cardiaques, car l'*explanandum* dépend de l'*explanans* au sens de Woodward, mais on ne peut pas expliquer téléologiquement les battements du cœur à partir de la production sonore de l'organisme. On ne peut donc pas lui attribuer la fonction correspondante.

Conclusions de la quatrième partie

La téléologie n'est pas une pensée précausale ni préscientifique qui s'appliquerait aux objets naturels à défaut d'une autre explication plus mature. Ce n'est pas un produit dérivé de la psychologie ni une forme invalide de la causalité. Au contraire, c'est un mode d'interprétation de la réalité à part entière, aussi primordial que la pensée physique et peut-être plus fondamental que la pensée intentionnelle ; c'est un outil mental inné, autonome et complémentaire des deux autres.

La pensée téléologique ne précède pas la pensée physique, car les deux sont disponibles très tôt au cours du développement cognitif, et elle n'implique pas l'attribution de causes finales ni rétrogrades, car elle ne raisonne pas en termes de relations causales (causes-conséquences), mais instrumentales (moyens-fins). Elle ne dérive pas non plus de la psychologie naïve ni de notre familiarité avec les artefacts, car il s'agit d'une faculté indépendante, peut-être même antérieure aux capacités d'attribution intentionnelle. D'après la théorie de l'attitude téléologique, c'est un système inférentiel d'interprétation des actions qui repose sur le principe d'action rationnelle, qui est sensible au contexte et qui ne dépend pas de la capacité à lire les pensées d'un agent (théorie de l'esprit), ni à lui attribuer des états mentaux (attitude intentionnelle), ni à simuler ses actions (par les neurones miroir).

À partir de l'analyse d'un cas très simple, on peut montrer que les explications causales et téléologiques sont différentes et complémentaires, que la pertinence d'employer l'une ou l'autre, ou les deux, dépend de l'objet et de la situation, et que les explications téléologiques n'ont pas nécessairement d'implications physiques ni métaphysiques inacceptables. De fait, elles sont largement employées dans les sciences physiques sous la forme de théories et de principes variationnels qui impliquent la maximisation ou la minimisation d'une grandeur. De manière analogue, les explications téléologiques en biologie prennent notamment la forme de théories et de modèles fondées sur des principes d'optimalité.

D'après la psychologie cognitive, le fondement de toutes les explications téléologiques serait le principe d'action rationnelle ou efficiente qui est un principe d'optimalité appliqué au rapport entre les moyens, les fins et les contraintes d'une action. On peut donc l'appliquer par exemple au

mouvement d'une figure géométrique sur un écran, au déplacement d'une pièce au jeu d'échecs ou aux phénomènes biologiques comme la bioluminescence de certains animaux et bactéries sans qu'il soit nécessaire de leur attribuer des états mentaux et quelles que soient par ailleurs les causes de la chose expliquée.

Cela étant, toutes les explications téléologiques ne se valent pas. À partir du principe d'action efficiente, on peut en effet montrer que certaines sont fausses ou inacceptables. Et bien qu'elles soient souvent employées à tort, notamment dans le cadre de théories préscientifiques (créationnisme, animisme) ou scientifiquement caduques (vitalisme, finalisme), ce n'est pas le type d'explication qui est en cause, mais son usage, car on peut en dire de même des explications causales (croyances superstitieuses, magiques et surnaturelles, effet cigogne). En ce qui concerne le reproche d'anthropocentrisme, il n'est pas spécifique à la téléologie ni à la biologie, et il constitue vraisemblablement un aspect fondamental et nécessaire de la pensée scientifique en général.

En biologie comme en physique, on peut prédire l'évolution d'un système en termes téléologiques à partir de principes variationnels. Ces prédictions sont compatibles avec une explication causale, mais elles ne dépendent pas d'une histoire ni d'un mécanisme causal particuliers. De plus, la téléologie entendue comme un mode de raisonnement à rebours — partant des fins/problèmes pour chercher les moyens/solutions — permet de réaliser des prédictions qui autrement seraient pratiquement impossibles. Par ailleurs, la classification fonctionnelle rend elle aussi possibles des inférences et des explications qui sont scientifiquement pertinentes.

Les explications téléofonctionnelles peuvent être considérées comme des explications authentiques non seulement du point de vue de la psychologie cognitive, mais aussi de l'épistémologie. En effet, on peut montrer que la théorie de l'explication (causale) de James Woodward est généralisable aux explications téléologiques (non causales) dans la mesure où elles permettent l'identification de relations de dépendance contrefactuelle invariante entre l'*explanandum* et l'*explanans*. De plus, conformément aux conclusions tirées de la psychologie cognitive, les invariants identifiés par les explications mécaniques et téléologiques sont différents, mais pas contradictoires, c'est-à-dire qu'il s'agit d'explications complémentaires.

Cinquième partie :
APPROCHES NON CAUSALES

Either "function" is defined in terms of causes, in which case there is nothing intrinsically functional about functions, they are just causes like any others. Or functions are defined in terms of the furtherance of a set of values that we hold—life, survival, reproduction, health—in which case they are observer relative.

John Searle, *The construction of social reality* (1995, p. 16)

The attributes which go along with meaningful use of the phrase "the good of *X*", may be called *biological* in a broad sense. […] What I mean by calling the terms "biological" is that they are used as attributes of beings, of which it is meaningful to say they have a *life*. The question "What kinds or species of things have a good?" is therefore broadly identical to the question "What kinds or species of being have a life?".

Georg H. von Wright, *The varieties of goodness* (1963, p. 50)

Dans un tel produit de la nature toute partie, tout de même qu'elle n'existe que par toutes les autres, est aussi conçue comme existant pour les autres parties et pour le tout, c'est-à-dire en tant qu'instrument (organe); […] on la conçoit donc comme un organe produisant les autres parties (et en conséquence chaque partie comme produisant les autres et réciproquement); aucun instrument de l'art ne peut être tel, mais seulement ceux de la nature, qui fournit toute la matière nécessaire aux instruments (même à ceux de l'art); ce n'est qu'alors et pour cette raison seulement qu'un tel produit, en tant qu'être organisé et s'organisant lui-même, peut être appelé une fin naturelle.

Immanuel Kant, *Critique de la faculter de juger* (1790, § 65)

Introduction de la cinquième partie

Nous avons consacré la première partie de ce travail aux principales stratégies naturalistes de définition des fonctions biologiques, c'est-à-dire à celles ayant eu le plus d'influence dans le débat depuis les années '70. Plusieurs autres stratégies ont été proposées qui ont rencontré moins de succès ou sont trop récentes pour avoir eu le temps de s'imposer. Parmi elles, certaines justifient l'attribution de fins à des systèmes naturels. Et, bien qu'elles ne rejetent pas le naturalisme en tant que tel, elles s'écartent du « naturalisme étroit » qui impose certaines restrictions métaphysiques quant au type d'entités et de propriétés qui peuvent légitimement faire partie du monde « tel qu'il existe véritablement ». Nous allons donc consacrer la dernière partie de ce travail à examiner trois de ces approches : mentaliste, valorative et organisationnelle.

Contrairement aux précédentes, ces approches ne permettent pas d'interpréter les explications téléofonctionnelles comme des explications causales déguisées, car elles impliquent l'attribution d'états mentaux ou de valeurs naturelles, ou un régime causal spécifique aux êtres vivants. Toutes trois permettent de soulever deux questions demeurées latentes tout au long de ce travail et que nous n'avons abordées que dans la seconde partie à propos du vivant. Il s'agit des questions portant sur la réalité et l'objectivité des attributions téléofonctionnelles. Dans quelle mesure peut-on dire que les fonctions et les fins, les valeurs et la normativité, ou encore les intentions, existent réellement dans la nature ? Au contraire, si elles ne sont pas dans la nature mais dans le regard, c'est-à-dire si elles sont relatives à l'observateur, alors dans quelle mesure les attributions correspondantes peuvent-elles néanmoins être objectives ? Nous verrons que différentes versions de ces approches apportent des réponses différentes à ces questions.

Les approches mentaliste, valorative et organisationnelle défendent toutes trois des positions assez proches de la nôtre. Nous allons donc nous servir de cette similarité comme marche-pied pour développer et nuancer les thèses que nous défendons. Les chapitres respectifs qui leur sont consacrés nous permettront ainsi de tirer quelques conclusions importantes pour clarifier notre posture.

Notre but n'étant pas de faire une révision de la littérature, nous n'allons pas examiner ici toutes les versions de ces trois approches ni mentionner tous les auteurs les ayant défendues. Pour la même raison, nous devons laisser de côté plusieurs autres approches non causales de l'explication téléofonctionnelle. En particulier, celles inspirées respectivement par Aristote et par Kant. Les premières ont été défendues notamment par Achinstein (1977), Ariew (2002), Barahona (2004), Cameron (2000), Marcos (2009) et Maund (2000). Les secondes par Cohen (2007), Ginsborg (2006, 2014), Perret (2015), Ratcliffe (2000), et Weber & Varela (2002). Les unes et les autres ne sont pas pour autant absentes dans les pages qui suivent, car il y a des chevauchements entre les conceptions aristotéliciennes ou kantiennes, d'un côté, et mentalistes, valoratives et organisationnelles, de l'autre. Plusieurs des auteurs que nous allons examiner ici défendent en effet des postures pouvant être classées aussi bien parmi les premières que parmi les secondes.

Chapitre XII

Approche mentaliste

L'une des principales approches non naturalistes de la téléologie biologique est le mentalisme ou téléomentalisme. Elle consiste à dire que les comportements dirigés vers un but ont pour modèle le comportement intentionnel humain et ne s'appliquent *stricto sensu* qu'aux êtres dotés d'un esprit, de sorte que l'attribution d'une finalité au comportement des êtres vivants et des machines relève dans la plupart des cas de l'analogie ou de la métaphore. Elle a été défendue, sous des formes différentes, par Georg H. von Wright, Curt J. Ducasse, Andrew Woodfield, Scott Sehon, Lowell Nissen, Richard Dawkins, Daniel Dennett et d'autres.

Nous verrons que certaines interprétations du mentalisme peuvent servir à appuyer et à justifier, depuis une perspective philosophique, les conclusions que nous avons tirées à partir des études empiriques en psychologie cognitive.

1. Téléomentalisme strict

D'après Lowell Nissen (1997), le langage téléologique implique nécessairement un agent intentionnel : l'explication téléologique du comportement d'un animal n'est justifiée que lorsque celui-ci possède une vie mentale suffisante, bien qu'elle ne soit ni consciente ni verbale. Dans tous les autres cas, le langage téléologique implique une intentionnalité extérieure à l'organisme. Le comportement des organismes inférieurs et les fonctions biologiques doivent ainsi être analysés de la même manière que les artefacts. Dès lors, il faut soit admettre dans notre conception du monde l'existence d'un agent intentionnel surnaturel, soit exclure des sciences de la vie l'usage littéral de la téléologie. Cette conception conduit à penser que la biologie pourrait se passer de la téléologie dans la plupart des cas.

D'après C.J. Ducasse (1925), seuls les actes des entités capables de désirs et de croyances sont en mesure d'être téléologiques, mais peu importe que les mots « désir » et « croyance » soient interprétés en termes de conscience ou en termes de neurones et d'influx nerveux, c'est-à-dire en termes mécaniques ; ce qui compte, c'est qu'il soit *vrai* que des désirs et des croyances sont présents (quelle que soit la façon dont on les décrit) pour qu'il y ait téléologie. Cela étant, lorsque l'explication d'un événement implique des désirs et des croyances, c'est-à-dire lorsque la cause de cet événement est bien l'effet d'une intention, alors non seulement cette explication est téléologique, mais c'est là la seule explication correcte. Autrement dit, bien que les désirs et les croyances puissent être décrits en termes mécaniques, les actions intentionnelles ne sont explicables qu'en termes téléologiques.[187]

2. États internes et analogies

L'une des critiques les plus courantes contre le téléomentalisme consiste à dire que beaucoup de systèmes dont le comportement semble dirigé vers un but n'ont vraisemblablement ni désirs ni croyances (Bedau, 1990, 1992b; Boorse, 2002; Collins, 1978; Nagel, 1977b). Une réponse possible à cette critique consiste à faire une lecture analogique ou métaphorique de la téléologie mentaliste tout en prenant au sérieux le discours téléologique, ce que ne fait pas Nissen. Quand on dit par exemple qu'un animal court pour échapper à un prédateur, cela n'implique pas nécessairement qu'il ait des états mentaux, c'est-à-dire des désirs et des croyances au sens où nous l'entendons chez l'humain, mais il nous semble néanmoins que son comportement trahit tout de même quelque chose qui est analogue au désir de sauver sa vie et à la croyance qu'en courant très vite il y arrivera. De même, quand on observe certains animaux qui emploient des outils et qui construisent des artefacts, comme les ruches des abeilles, les nids des oiseaux, les barrages des castors, il nous semble que leur comportement est analogue à celui des artisans humains, bien qu'ils n'aient sans doute pas les mêmes états mentaux.

Les comportements dirigés vers un but sont ce qui a motivé la conception cybernétique de la téléologie, en partant de l'idée que le but d'un système n'est pas déterminé par les intentions de son créateur ni par le choix d'un observateur, mais par sa structure interne. On se souvient par exemple que pour Nagel le fait d'être dirigé vers un but était une propriété du système en vertu de l'organisation de ses parties et que l'on pouvait reconnaître ce but sans avoir à analyser sa structure interne grâce aux mécanismes de compensation que révèle son comportement. Mais la

187 Pour un développement postérieur et plus élaboré de la conception de cet auteur, voir Ducasse (1959).

cybernétique s'appuie sur le modèle de la machine et, tout en légitimant l'usage scientifique et non intentionnel de notions d'origine psychologique, comme celles de but (*goal*), de fin (*end*) et de dessein (*purpose*), elle les réduit à des mécanismes de rétroaction ou rétro-alimentation (*feed-back*). L'approche cybernétique permet donc d'expliquer un comportement non pas seulement à partir de ses causes, comme dans la démarche mécaniste classique, mais aussi de ses effets attendus, c'est-à-dire en termes de buts, sauf que les buts en question ont eux-mêmes une traduction matérielle, comme la température de référence d'un thermostat ou la cible d'un missile autoguidé.

Andrew Woodfield (1976) partage avec la cybernétique l'idée que le but d'un système est déterminé par sa structure interne, mais il adopte une démarche différente. Au lieu d'analyser les comportements téléologiques en termes mécaniques (mécanismes de rétroaction), il le fait en termes psychologiques. Il considère ainsi que la caractéristique essentielle des membres de plein droit de la classe des systèmes dirigés vers un but est leur capacité à se trouver dans un certain type d'état interne qui est analogue à un état mental (1976, p. 201).

L'objectif poursuivi par cet auteur n'est pas non plus le même que celui des partisans de la cybernétique. Nagel cherche à objectiver les mécanismes causaux responsables du comportement « mystérieux » attribué aux systèmes biologiques. Il veut connaître les conditions de possibilité des processus dynamiques auto-régulés dirigés vers une fin. Woodfield cherche quant à lui à rendre intelligible le langage téléologique. Sa préoccupation principale n'est pas d'opposer au vitaliste une explication scientifique des comportements téléologiques, mais plutôt de rendre explicite la structure grammaticale commune aux différentes descriptions téléologiques et en examiner les conditions de vérité. La téléologie, dit-il, n'est pas quelque chose dont les scientifiques pourraient montrer qu'elle n'existe pas, car elle existe dans la mesure même où les descriptions téléologiques (TDs, *Teleological Descriptions*) peuvent être vraies. Le travail du philosophe consiste à réaliser un travail d'analyse et de clarification conceptuelle qui permette de dégager ces conditions de vérité.

Or, d'après cette analyse, le concept de but — quand on parle de système dirigés vers un but — est essentiellement mentaliste. Il s'applique d'abord aux êtres humains et aux animaux dits « supérieurs » auxquels on reconnaît des états mentaux correspondants, à savoir des désirs et des croyances. Et c'est seulement par analogie que l'on étend le concept aux animaux « inférieurs » et à des artefacts comme le missile auto-guidé. Cependant, une description téléologique appliquée à un organisme trop simple pour avoir des états mentaux peut néanmoins être vraie s'il dispose d'états internes analogues à des désirs et des croyances, mais c'est ensuite aux scientifiques de déterminer quelles structures et quels processus internes sont susceptibles de le vérifier. Ce n'est pas au philosophe

analytique de fixer les critères d'appartenance à la catégorie des systèmes dirigés vers un but, d'autant moins que ces critères sont susceptibles d'évoluer au gré des découvertes scientifiques et des progrès techniques.

Dans une certaine mesure, les deux approches sont complémentaires : l'analyse cybernétique du comportement des systèmes dirigés vers un but permet de mieux comprendre pourquoi les biologistes ont tendance à en donner une description téléologique ; l'analyse conceptuelle de ces descriptions permet de mieux saisir leur structure commune et leurs conditions de vérité. Elles sont complémentaires parce que, face à un même phénomène — l'apparence de finalité dans le comportement de systèmes réputés non-intentionnels, — Woodfield et les naturalistes donnent des réponses différentes à des questions différentes. Tandis que Nagel et Collins cherchent une explication de type scientifique qui rende compte des comportements dirigés vers un but, Woodfield cherche à rendre compte de l'attribution de buts à ces systèmes. C'est-à-dire qu'il cherche à comprendre *notre* comportement linguistique vis-à-vis du comportement de ces systèmes. Les uns et les autres sont d'accord pour reconnaître que la nature téléologique d'un système est liée à sa structure interne (laquelle est responsable de son comportement externe), mais ils en donnent une interprétation différente, de même qu'ils donnent une signification différente à la notion de « but » avec laquelle ils décrivent ce comportement.

Une autre différence entre Woodfield et ses détracteurs tient à leur manière respective de concevoir la finalité. Nagel (1977b) et Collins (1978) cherchent à identifier des propriétés de ces systèmes et ils exigent la même chose de Woodfield. Mais cette lecture ignore un aspect essentiel de l'analyse mentaliste : les buts que nous attribuons aux systèmes non-intentionnels sont des métaphores.

3. Métaphores et objectivité

Le mentalisme de Woodfield consiste à dire que le paradigme du comportement dirigé vers un but est le comportement humain et que son attribution à des animaux et à des machines relève de l'analogie ou de la métaphore. Mais à force d'avoir été utilisée une métaphore peut mourir, de sorte que l'on finit par la prendre à la lettre. C'est ainsi que l'on va métaphoriquement attribuer un but aux machines pour s'étonner ensuite — l'origine métaphorique de cette attribution étant oubliée — qu'une machine puisse avoir un but. À partir de ce constat, la question de savoir comment, causalement, un système mécanique ou biologique peut manifester un comportement dirigé vers un but deviendrait secondaire si, après tout, il ne s'agissait que d'une manière de *décrire* ce comportement — mais ce n'est pas tout à fait la position de Woodfield. Depuis cette perspective, l'aura de mystère qui entoure habituellement la téléologie

biologique, dont les vitalistes tirent profit et dont les naturalistes cherchent à se débarrasser, n'aurait tout simplement pas lieu d'être. La finalité ne serait pas quelque chose (une propriété) que ces systèmes *ont* et qui réclame explication, mais quelque chose (un mode de description) qu'on leur *prête*.

Ce dernier point pose directement la question de l'objectivité des descriptions téléologiques. Certains considèrent, dit Woodfield, que les TDs ne sont pas des énoncés empiriques et que, par conséquent, elles ne sont ni vraies ni fausses, mais qu'elles expriment notre manière de voir les choses. La téléologie, dès lors, ne serait qu'une projection de l'esprit. Woodfield, au contraire, affirme que sa recherche des conditions de vérité des TDs repose sur le présupposé que celles-ci peuvent être objectivement vraies ou fausses (1976, p. 26). Il lui faut donc se démarquer de ce qu'il appelle le projectionnisme kantien :

> « The question to be settled is: Where is teleology really located – in reality, in language, or in the mind? I have presumed that the presence of teleological connectives in descriptions is to be justified by reference to the presence of objective features in what those descriptions describe. The alternative view, which I shall call *projectionism*, is that TDs project on to things a 'property' which the things do not really possess. The average man does not know this. He thinks he is saying something about things when he uses teleological sentences. According to projectionism, this is an illusion. What he is really doing is metaphorically projecting his own teleological attitudes on to the world. He is saying, in effect, 'It looks at me as if an intelligence has been at work here'. His utterance masquerades as a categorical assertion about reality, but really it expresses his own state of mind. Probably the most illustrious proponent of a theory of natural teleology which is basically projectionist was Kant. It is crucial for me to face up projectionism, because my search for truth-conditions would be futile if TDs couldn't ever be true. » (Woodfield, 1976, p. 25-26)

L'auteur veut concilier l'objectivité et la vérité des descriptions téléologiques (TDs) avec leur caractère métaphorique. Dans le passage cité, il oppose deux interprétations de la téléologie. Selon la première, les TDs sont l'expression subjective d'un « état d'esprit » projeté sur le monde et ne reflètent qu'une illusion de réalité. L'autre, au contraire, soutient que les TDs se rapportent aux choses elles-mêmes et permettent de les catégoriser. Mais si, comme il l'affirme par ailleurs, les TDs sont analogiques et métaphoriques, alors les métaphores devraient être pourvues, selon lui, d'un pouvoir de catégorisation du réel.

On ne trouve pas dans le texte de l'auteur une formulation explicite de cette idée, mais certains passages évoquent un modèle de catégorisation prototypique comme celui développé à cette époque par Eleanor Rosch (1973, 1975). En effet, Woodfield construit la catégorie « *goal-di-*

rected » à partir d'un cas central, l'être humain, autour duquel s'organisent graduellement d'autres systèmes par ordre de « proximité analogique » ou de similarité au cas central (1976, p. 202).

La catégorie n'a pas de limites bien définies et ses critères d'appartenance peuvent varier. Le fait d'être vivant a été, dit-il, un trait essentiel de la catégorie et un critère d'exclusion pour les systèmes mécaniques. Ces derniers n'étaient pas reconnus comme membres de plein droit de la catégorie des systèmes dirigés vers un but et ne s'y voyaient associés qu'en un sens métaphorique. Ils se trouvaient à la périphérie de la catégorie sans y appartenir. Mais, la métaphore étant morte, les critères d'appartenance ont glissé et les servomécanismes y ont trouvé leur place.

Le caractère métaphorique et prototypique de la catégorisation est une des raisons pour lesquelles Woodfield ne peut pas fournir de règles claires pour décider quand, dans un système mécanique ou chez le têtard, un état interne est suffisamment similaire à un désir ou à une croyance pour que le système soit considéré comme « ayant un but ». Les catégories construites à partir de prototypes n'ont pas de conditions nécessaires et suffisantes d'appartenance et leurs membres ne sont pas tous équivalents mais plus ou moins représentatifs de la catégorie. Cela étant, dans la théorie de Rosch on peut identifier des attributs qui (à un moment donné et dans une communauté linguistique donnée) jouent un rôle central dans l'organisation d'une catégorie. Par exemple, avoir des plumes, un bec et la capacité de voler sont des caractéristiques « essentielles » — mais pas suffisantes ni absolument nécessaires — de la catégorie « oiseau ». Pour les systèmes dirigés vers un but, la caractéristique essentielle que leur reconnaît Woodfield est la capacité à agir en fonction de désirs et de croyances, et cette caractéristique est ce qui rend objectives les descriptions téléologiques.

Le sens qu'il donne à cette objectivité n'est pas très clair. Signifie-t-elle que l'intentionnalité d'un robot est un fait indépendant des jugements que nous portons sur lui ? Probablement pas, puisque la catégorisation des systèmes en tant que *goal-directed* est sujette à l'évolution des pratiques linguistiques (perte de valeur métaphorique des TDs) et à l'évolution des jugements (altération des règles d'inclusion de la catégorie pour y associer des membres « inanimés »). En revanche, cela signifie peut-être que le critère intersubjectif d'inclusion dans cette catégorie est, lui, un fait objectif — c'est-à-dire qu'il est empiriquement observable, indépendant du jugement, etc. De même que le fait d'avoir des plumes est un trait objectif servant de critère d'appartenance à la catégorie « oiseau », la possession d'un certain état interne servirait de critère objectif d'appartenance à la catégorie « *goal-directed* ».

L'auteur semble reconnaître à certaines machines une intentionnalité véritable, objective, en vertu de leurs états internes. Mais il affirme simultanément que cette reconnaissance est le fruit d'un glissement sémantique : dans la mesure où la description téléologique du comporte-

ment d'un robot a perdu sa valeur métaphorique, on ne peut plus se contenter de faire *comme si* il possédait un point de vue et des états intentionnels ; on doit prendre, dit-il, la description téléologique à la lettre.

Une manière de résoudre la contradiction apparente entre ces différentes idées consiste à distinguer, d'une part, l'appartenance à une catégorie et, d'autre part, le statut de cette catégorie. L'appartenance peut être objective quand bien même la catégorie serait arbitraire. Par exemple, dans le système monétaire, chaque pièce de monnaie est soit vraie soit fausse de manière objective bien que la catégorie « fausse monnaie » soit une convention arbitraire qui ne correspond à rien dans la réalité dite « objective » (car elle est entièrement relative à nos pratiques et représentations sociales). Au risque d'en forcer l'interprétation, on pourrait comprendre la posture de Woodfield de manière analogue : la catégorie des systèmes dirigés vers un but est relative à nos pratiques linguistiques et à nos représentations, mais n'importe quel objet (naturel ou artificiel) appartient ou n'appartient pas, de façon objective, à cette catégorie.

La réflexion menée dans la deuxième partie de cette thèse sur les stratégies de définition de la vie nous avait conduits à des conclusions similaires. Bien que nous puissions nous mettre tous d'accord pour reconnaître que l'appartenance à la catégorie des êtres vivants est un fait objectif, c'est-à-dire que n'importe quel objet physique (lapin, caillou, virus) est vivant ou ne l'est pas, cette appartenance dépend de la conception que l'on adopte. Or, contrairement au concept de planète, pour lequel les astronomes ont adopté une définition stipulative et conventionnelle, les biologistes ne s'accordent pas sur la conception ni sur la définition des concepts de vie et de vivant. Cependant, contrairement à la monnaie et aux planètes, l'appartenance d'un objet à la catégorie du vivant n'est pas arbitraire ni strictement conventionnelle. Pourtant, nous avons dit que ce ne sont pas des genres naturels et que les limites de ces catégories ne se trouvent pas dans la nature, mais dans le regard ou dans le discours que l'on porte sur eux. Et nous avons recouru pour expliquer cela à la distinction qu'établit Dennett entre les propriétés « suspectes » et « charmantes » : la vie est charmante, les planètes et la monnaie sont suspectes (CHAP. V, SECT. 6.3).

Si l'on interprète la posture de Woodfield à la lumière de cette reflexion, on dira que le fait d'être dirigé vers un but est une propriété charmante. D'un côté, elle dépend des caractéristiques objectives des systèmes eux-mêmes, c'est-à-dire de leurs comportements et de leurs états internes. De l'autre, elle dépend de la subjectivité des observateurs, c'est-à-dire de la façon dont ils voient et décrivent ces systèmes, leurs comportements et leurs états. Cette double dépendance définit les conditions de vérité des descriptions téléologiques : un système est (véritablement) dirigé vers un but, selon cette interprétation, s'il possède (objectivement) certains états internes qui, pour une classe donnée d'observateurs, sont perçus (subjectivement) comme analogues à des désirs et des croyances.

Dans la mesure où cette perception est mouvante, les frontières de la catégorie le sont aussi, mais elles ne sont pas pour autant arbitraires ni conventionnelles.

En conclusion, on peut dire que l'approche mentaliste de Woodfield justifie les descriptions téléologiques en montrant qu'elles peuvent être à la fois vraies et objectives dans la mesure où elles se rapportent à quelque chose qui est prétendument vérifiable et objectivable, à savoir la possession d'états internes analogues à des désirs et des croyances. C'est une justification qui s'appuie sur la psychologie : puisqu'il est légitime de décrire un comportement intentionnel humain en termes téléologiques, il doit l'être aussi de le faire pour des comportements non humains qui remplissent des conditions analogues d'intentionnalité. C'est du moins de cette manière, dit Woodfield, que fonctionne notre langage.

L'un des problèmes de cette justification, c'est qu'elle est incomplète. Elle ne s'applique en effet qu'à certaines descriptions téléologiques, à savoir celles qui portent sur le comportement des agents et celles qui portent sur les fonctions des artefacts. Et nous ne savons pas au juste quels sont les systèmes auxquels elle s'applique, c'est-à-dire lesquels ont véritablement des états analogues à des croyances. Pour ce qui est des fonctions biologiques et du comportement des organismes dits inférieurs, l'analyse de l'auteur est différente, comme nous le verrons au chapitre suivant.

Un autre problème de cette justification, de notre point de vue, est le rapport de subordination qu'elle établit entre la téléologie et la psychologie. Nous avons vu que ce sont des choses vraisemblablement indépendantes au niveau du développement cognitif et que, si l'on en croit la théorie de Gergely et Csibra, la pensée téléologique n'a pas besoin de la pensée intentionnelle pour se justifier. Woodfield s'est appuyé sur une analyse philosophique du langage pour en dégager la structure du discours téléologique, mais la psychologie cognitive nous permet aujourd'hui d'aller plus loin dans cette direction.

4. Réalisme et interprétationnisme

Comment le fait de désirer ou de croire, quand on parle d'un système mécanique ou biologique, peut-il être à la fois objectif et dépendant de nos représentations ? On ne trouvera pas de réponse claire à cette questions dans le texte de Woodfield, mais Daniel Dennett développe une intéressante réflexion en ce sens dans *The Intentional Stance*. L'auteur s'y intéresse particulièrement à l'attribution de croyances. Il y défend l'idée apparemment contradictoire selon laquelle les croyances seraient des phénomènes à la fois objectifs et relatifs au point de vue de celui qui les attribue, c'est-à-dire relatifs à une stratégie prédictive :

> « My thesis will be that while belief is a perfectly objective phenomenon (that apparently makes me a realist), it can be discerned only from the point of view of one who adopts a certain *predictive strategy*, and its existence can be confirmed only by an assessment of the success of that strategy (that apparently makes me an interpretationist). » (Dennett, 1987, p. 15)

Dans ce passage, l'auteur situe sa propre position entre le réalisme et l'interprétationnisme. Le réalisme considère possible, en principe, de confirmer une attribution de croyance en cherchant « dans la tête du croyant » quelque chose qui lui corresponde. C'est-à-dire que le fait d'avoir ou pas une certaine croyance dépendrait en dernière instance d'un fait objectif interne, comme un état physique du cerveau par exemple. Ainsi, avec des connaissances plus approfondies en psychologie physiologique, nous pourrions être capables, en principe, de déterminer les états de croyance d'un individu en observant les états internes de son cerveau. Le fait d'avoir une croyance, dit-il, est comparable aux yeux d'un réaliste au fait d'être infecté par un virus. Aux yeux d'un interprétationniste, en revanche, le fait d'avoir une croyance est plutôt comparable au fait d'avoir du talent, d'avoir du goût ou d'être quelqu'un de bien. La réponse obtenue dépend de la personne à qui l'on pose la question et du point de vue qu'elle adopte. C'est une affaire d'interprétation.

Pour prédire le comportement d'un objet, dit Dennett, on peut adopter plusieurs attitudes ou stratégies interprétatives. L'une d'elles consiste à le considérer comme un agent rationnel et lui attribuer des croyances et d'autres états mentaux à contenu représentationnel. Cette stratégie, dite intentionnelle, est efficace avec les êtres humains et certains animaux ; elle l'est moins avec les objets inertes. Pour prédire le comportement de ces derniers, il est généralement préférable de les considérer comme des systèmes physiques régi par les lois de la physique, plutôt que comme des systèmes intentionnels mûs par des états mentaux. A l'inverse, pour prédire le comportement d'un être vivant, le considérer comme un simple objet physique régi seulement par les lois de la physique est souvent inefficace ou totalement inutile. La stratégie intentionnelle aura donc plus ou moins de succès prédictif suivant ce à quoi on attribue des croyances.

On peut raisonnablement supposer que le succès ou l'échec de cette stratégie prédictive est relatif à la nature — intentionnelle ou pas — des objets auxquels on l'applique. Pour un réaliste, cela signifie que les objets ont une nature intrinsèque indépendante de la stratégie adoptée qui en explique le succès ou l'échec. Pour Daniel Dennett, au contraire, cela signifie que la nature intentionnelle de l'objet dépend directement de l'issue de la stratégie qu'on lui applique : un système intentionnel est un système dont l'attitude intentionnelle permet de prédire avec succès le comportement :

> « I will argue that any object—or as I shall say, any system—whose behavior is well predicted by this [intentional] strategy is in the fullest sense of the word a believer. *What it is* to be a true believer is to be an *intentional system*, a system whose behavior is reliably and voluminously predictable via the intentional strategy. » (Dennett, 1987, p. 15)

On retrouve chez Dennett comme chez Woodfield une distinction marquée entre les systèmes qui possèdent *réellement* des croyances et des désirs et les systèmes que l'on peut traiter métaphoriquement *comme si* ils en avaient. Le critère de démarcation se situe, pour l'un, dans les états internes du système, et, pour l'autre, dans le succès de la stratégie interprétative. Tous deux affirment que leur critère, bien que relatif à nos pratiques et représentations, est néanmoins objectif, de sorte que l'objectivité dont ils parlent n'est pas une indépendance stricte à l'égard de l'« observateur » au sens du réalisme. En effet, si l'appartenance d'un objet à la catégorie des systèmes intentionnels n'est pas fixée *sub specie aeternitatis,* elle n'est pas non plus choisie arbitrairement par l'« observateur ».

Si le succès ou l'échec d'une stratégie prédictive est relatif à la nature de l'objet auquel elle se rapporte, alors il doit bien y avoir quelque chose dans l'objet lui-même, c'est-à-dire dans le monde, qui détermine le succès ou l'échec des prédictions. En effet, là où il n'y a que du hasard, rien n'est prédictible, dit Dennett dans un texte postérieur (Dennett, 1991, p. 30). Le succès d'une prédiction dépend donc de l'existence dans le monde d'un certain ordre ou *pattern* que l'on peut exploiter. Mais où se trouve ce *pattern* ? Est-il dans la structure interne des objets, par exemple sous la forme de structures cérébrales ? Ou existe-t-il plutôt dans le comportement observable des agents lorsqu'on le soumet à une interprétation radicale depuis l'attitude intentionnelle ?

Nous avons vu que Woodfield, Ducasse et Lowell, chacun à sa façon, défendent la première option, tandis que Dennett penche pour la seconde. Les deux ne sont pas incompatibles, bien au contraire, puisque le comportement externe d'un agent dépend en grande partie de ses structures internes, de sorte que ceux qui possèdent réellement des états mentaux intentionnels manifestent souvent des comportements intentionnels. Mais l'argument de Dennett va plus loin. D'après lui, nous n'avons pas besoin de regarder dans le cerveau de quelqu'un pour savoir s'il est véritablement intentionnel, l'observation de son comportement étant suffisante. Et ce n'est pas parce que le comportement externe trahit la structure interne, mais plutôt parce que le comportement intentionnel est ce qui *définit* un agent intentionnel — quelles que soient ses structures internes. Sauf que le caractère intentionnel du comportement dépend de l'interprétation ou de la stratégie prédictive que l'on adopte.

C'est là l'une des difficultés principales de la posture de Dennett. Comment l'intentionnalité peut-elle être objectivement présente dans le comportement d'un agent et nonobstant relative à une interprétation ? Les exemples et développements de l'auteur étant trop longs à résumer ici (voir Dennett, 1991), nous prendrons un autre exemple, plus simple, au risque de simplifier aussi son raisonnement. Considérons une série incomplète de nombres comme la suivante :

(S_1) … 2, 4, 6, 8, …

Y a-t-il dans cette série un ordre quelconque ? Les notions d'ordre et de hasard ont une définition mathématique précise en Théorie algorithmique de l'information que l'on peut exprimer en termes de compression informatique : un objet est ordonné s'il est compressible, c'est-à-dire s'il existe un algorithme qui le décrive ou le produise avec une quantité d'information moindre que celle de l'objet lui-même. Par exemple, il existe un algorithme très simple pour produire l'ensemble des nombres pairs, à tel point que n'importe quel enfant pourrait en réciter la liste ; cela veut dire que la série infinie des nombres pairs est extrêmement ordonnée. Il existe aussi des algorithmes pour calculer les décimales du nombre π, ce qui veut dire qu'elles présentent un ordre intrinsèque malgré leur caractère apparemment aléatoire. Ce qui est réellement aléatoire, selon la théorie algorithmique de l'information, c'est ce qui est incompressible, c'est-à-dire ce qui ne peut pas être décrit plus simplement.

Y a-t-il donc dans la série S_1 un ordre quelconque ? Oui, car il s'agit de la liste des nombres pairs de 2 à 8. Et le fait de connaître cet ordre nous permet d'en prédire facilement les éléments suivants.

Considérons maintenant la série numérique suivante :

(S_2) … **1**, **2**, **3**, **5**, …

Il est facile d'y reconnaître le début de la suite des nombres entiers naturels auquel manquerait le 4. S'il en est ainsi, alors il existe effectivement un ordre intrinsèque dans la série S_2 qui permet de la décrire et de la compléter :

(S_3) **1**, **2**, **3**, 4, **5**, 6, 7, …

Cependant, on peut aussi reconnaître dans S_2 le début de la suite de Fibonacci auquel manquerait le premier élément :

(S_4) 1, **1**, **2**, **3**, **5**, 8, 13, …

De même, on peut reconnaître dans S_2 le début de la suite des nombres premiers auquel s'ajouterait le 1 qui n'en fait pas partie :

(S_5) ~~**1**~~, **2**, **3**, **5**, 7, 11, 13, …

Ou bien encore le début de la suite des nombres impairs auquel s'ajouterait le 2 qui n'en fait pas partie :

(S_6) **1**, ~~**2**~~, **3**, **5**, 7, 9, 11, …

La suite de Fibonacci, celle des nombres premiers, celle des nombres pairs et celle des nombres entiers sont quatre interprétations de la même série. Les quatre sont approximatives en ce qu'elles ont un élément de moins ou de trop. Les quatre correspondent à un algorithme extrêmement simple permettant de décrire et de construire une suite numérique infinie. Du point de vue de la théorie algorithmique de l'information, chacune de ces interprétations décrit donc un ordre intrinsèque objectivement présent dans la série S_2. Pourtant, l'ordre décrit n'est pas le même. Elles décrivent des *patterns* bien réels, mais ces *patterns* sont relatifs à une interprétation et ces interprétations sont ici mutuellement incompatibles.

Un certain nombre de philosophes diraient que pour savoir laquelle de ces interprétations est correcte et lequel de ces *patterns* est réel, il faut connaître les processus internes ayant produit S_2, à savoir l'algorithme de la machine ou les états mentaux de la personne. L'argument de Dennett consiste en partie à montrer que le recours à une structure sous-jacente n'est pas toujours utile, car dans certains cas la connaissance détaillée de cette structure ne permet pas de lever l'ambiguïté. Dans ces cas là, on ne peut que recourir au succès prédictif pour trancher entre plusieurs interprétations rivales. Et encore n'est-ce pas suffisant puisque deux interprétations pourraient être aussi bonnes l'une que l'autre quant à la concision de leur description et à la fiabilité à long terme de leurs prédictions, bien qu'elles ne soient pas toujours d'accord l'une avec l'autre et qu'elles se trompent parfois sur des points importants.

Le raisonnement de Dennett porte sur l'interprétation du comportement humain et, plus généralement, sur l'intentionnalité et l'attribution à autrui de croyances, de désirs et autres états mentaux représentationnels. Il défend à la fois la réalité des attributions, y compris en cas d'interprétations rivales, et leur indétermination radicale qui fait que le choix d'une interprétation ou d'une autre dépend de l'observateur sans qu'il y ait un fait plus profond (un *deeper fact*), à savoir une réalité sous-jacente, qui permette de trancher la question[188] :

> « I see that there could be two different systems of belief attribution to an individual which differed *substantially* in what they attributed—even in yielding substantially different predictions of the individual's future behaviour—and yet where no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not. In other words, there could be two different, but equally real, patterns discernible in the noisy world. The rival theorists would not even agree on which parts of the world

188 À propos de cette question, voir la critique de Millikan (2000).

> settle the issue. The choice of a pattern would indeed be up to the observer, a matter to be decided on idiosyncratic pragmatic grounds. » (Dennett, 1991, p. 49)

De cette manière, Dennett semble renoncer, du moins en ce qui concerne l'intentionnalité, à la conception de la vérité comme correspondance entre le langage ou la pensée (ici l'interprétation ou l'attitude intentionnelle) et une prétendue réalité mentale indépendante de l'observateur dont le siège serait, par exemple, dans les structures cérébrales. Cela ne veut pas dire pour autant qu'il renonce à une conception substantive de la vérité ni qu'il embrasse une conception postmoderne et relativiste (voir à ce propos Dennett, 2000). Bien au contraire, si les systèmes intentionnels sont le produit de la sélection naturelle, c'est bien parce qu'il existe une différence entre la vérité et l'erreur ou entre la réalité et l'apparence, car la nature sanctionne sévèrement les fautes d'interprétation et l'échec des prédictions d'un organisme dans la relation qu'il entretient avec ses prédateurs, ses proies et ses congénères (voir Dennett, 1987, Chapitre 8, 1996). C'est la réalité qui détermine le succès ou l'échec de nos prédictions. Réciproquement, si la psychologie populaire (*folk psychology*) nous rend capables de comprendre et d'anticiper le comportement d'autrui en lui attribuant des états mentaux, alors il doit y avoir quelque chose de vrai et de réel dans ces attributions (Dennett, 1991).

5. Téléologie intrinsèque et extrinsèque

Le raisonnement de Dennett est applicable à la téléologie biologique. Cet auteur distingue trois attitudes interprétatives ou *stances* : physique, artefactuelle (*design*) et intentionnelle, mais nous avons vu précédemment que certains chercheurs en psychologie cognitive défendent l'existence d'une quatrième attitude, téléologique, permettant l'attribution de fins sans que celles-ci impliquent nécessairement l'attribution simultanée d'états mentaux (voir CHAP. VIII, SECT. 2.2, p. 278). On peut par exemple affirmer que la plante se tourne vers la fenêtre pour maximiser son ensoleillement sans pour autant supposer qu'elle ait une quelconque idée ou représentation mentale de la fenêtre, ni du soleil, ni de la fin qu'on lui attribue. Cependant, on peut difficilement nier qu'elle ait une certaine forme de perception de la lumière qui la pousse à s'orienter vers elle de sorte que la fin qu'on lui attribue se réalise. Mais comment comprendre cette attribution ? A-t-elle réellement des fins qui lui soient propres et intrinsèques, comme on pourrait le dire d'une personne ? A-t-elle réellement des fins propres et extrinsèques, comme celles d'un artefact ? Ou faisons-nous seulement *comme si* elle avait des fins, sans vraiment les prendre au sérieux ?

Jusqu'ici, nous avons soutenu que le langage téléologique et les attributions de fins et de fonctions ne sont pas seulement des façons de parler ni des fictions utiles. Nous écartons donc la troisième option. Avant de trancher entre les deux premières, qui opposent la téléologie intrinsèque et extrinsèque, nous devons examiner la façon dont le philosophe aborde la même question à propos de l'intentionnalité humaine.

Dans *The Intentional Stance*, Dennett signale que l'un des points d'achoppement de sa théorie pour bon nombre de ses collègues est la distinction qu'ils établissent — et que lui refuse — entre intentionnalités originale et dérivée, la première étant celle des humains et la seconde celle des artefacts. Pour lui, cette distinction n'a pas lieu d'être dans la mesure où nous-mêmes, êtres humains, sommes des artefacts issus d'un lent et long processus d'évolution par sélection naturelle. Si les robots, aussi sophistiqués soient-ils, n'ont d'intentionnalité que dérivée, alors la nôtre est du même type, car nous avons été conçus et fabriqués par Mère Nature ; et s'il devait y avoir une intentionnalité originale, ce serait la sienne (1987, p. 299, 1996, p. 53).

Pour arriver à ces conclusions, l'auteur applique l'attitude intentionnelle à la sélection naturelle et celle du *design* à ses productions (les plantes, les humains, etc.), mais on ne peut manquer de remarquer un certain flottement dans sa formulation. Tout en insistant sur le fait que Mère Nature est un horloger aveugle qui ne peut pas voir ni prévoir le futur (*no foresight*) et qui n'a pas non plus de buts ni de desseins (*no purpose*), il lui attribue cependant certaines des capacités de l'esprit humain et parle à son propos de « reconnaissance », d'« appréciations » et de « choix raisonnés » (c'est l'auteur qui met les guillemets), malgré son incapacité à se « représenter » ces choix et ces raisons. Pourtant, l'attitude intentionnelle consiste justement à interpréter et à prédire le comportement d'un agent en lui attribuant des états mentaux représentationnels. Or, nous avons vu plus haut que ces attributions doivent être prises au sérieux et pas seulement comme des métaphores ou des fictions utiles. Il y a donc une contradiction apparente dans ce discours que l'auteur lui-même n'hésite pas à reconnaître (1987, p. 314).

Pour lever la contradiction, il tâche alors de montrer que l'anthropomorphisme est non seulement utile, mais inévitable, à moins de renoncer à parler de fonctions biologiques, ce qui aurait un coût très élevé, car ce serait comme renoncer à parler de fonctions à propos des artefacts. Il ajoute que l'intentionnalité de Mère Nature est aussi réelle que la nôtre, mais il affirme simultanément que son intelligence est une illusion (contrairement à la nôtre) et qu'elle apprécie les raisons sans se les représenter. Il nous semble donc que l'auteur ne parvient pas à résoudre la tension entre le caractère mécanique et aveugle qu'il reconnaît à la sélection naturelle et l'anthropomorphisme auquel le conduit l'attitude intentionnelle. En effet, selon lui, il n'existe pas de position intermédiaire stable entre les deux (1987, p. 316) : soit on renonce au langage fonc-

tionnel, avec tout ce que cela implique, soit on accepte le principe que la *sélection* naturelle est bien nommée.

Pourtant l'attitude téléologique constitue bel et bien une position intermédiaire stable qui permet de relâcher la tension entre les extrêmes du mentaliste et du mécanisme. Elle permet de concevoir le processus d'évolution par sélection naturelle en termes de moyens et de fins (ou de solutions à des problèmes) sans qu'il soit nécessaire de parler ni de choix ni de représentation (avec ou sans guillemets).

Vermaas et al. (2013) proposent quant à eux de lever l'ambiguïté du *design stance* vis-à-vis de l'intentionnalité en y distinguant deux attitudes différentes. La première, qu'ils nomment « *teleological design stance* », est celle que l'on adopte face au comportement d'une entité que l'on décrit comme ayant une finalité et qui est composée de parties fonctionnelles. La seconde, dite « *intentional designer stance* », ajoute que la fin et les fonctions de cette entité lui ont été assignées par une autre que l'on décrit comme étant un agent rationnel.

> « In the teleological design stance a person *y* predicts the behaviour of an entity *x* by appeal to the assumption that *x* is an entity with a purpose and with parts that have functions that are all assigned by person *y*. In the intentional designer stance a person *y* predicts the behaviour of an entity *x* by appeal to the assumption that *x* is an entity with a purpose and with parts that have functions, *and* by appeal to the assumption that this purpose of *x* and the functions are assigned by an entity *z* that person *y* describes as a rational agent with certain overarching goals and certain perceptual and behavioural capacities. » (Vermaas et al., 2013, p. 1144)

La première est applicable aux organismes biologiques, à leurs organes et à leur comportement quand les chercheurs leur assignent des fins et des fonctions sur la base de l'observation ou de la théorie de l'évolution par sélection naturelle. La seconde est applicable au comportement des artefacts et des objets biologiques quand on suppose que ces derniers ont été conçus par « Mère Nature ».

L'attitude téléologique de Csibra et Gergely s'applique tout d'abord au comportement d'un agent auquel on attribue une fin. Elle n'implique pas de penser cet agent comme étant composé de parties fonctionnelles. Lorsque l'on attribue une fin au comportement du petit cercle qui se déplace vers le grand, peu importe en effet que le cercle soit simple ou composé. De même, lorsque l'on voit un simple point lumineux qui en poursuit un autre, l'interprétation téléologique de leur comportement repose sur leurs mouvements respectifs, lesquels ressemblent à ceux d'un prédateur et de sa proie. Le cercle et le point lumineux n'ont pas de *design*. Et lorsque nous attribuons une fin à la plante qui se tourne vers la fenêtre, nous n'avons pas non plus besoin de la concevoir comme un objet complexe. Il nous semble par conséquent que le *teleological design*

stance de Vermaas et ses collègues est une attitude interprétative plus exigente que le *teleological stance* des psychologues hongrois, car elle implique une compositionnalité et une structure fonctionnelle.

Il nous semble par ailleurs qu'elle est plus restreinte dans sa portée, car elle s'applique aux produits du processus de sélection naturelle ou d'un autre processus, mais pas au processus lui-*même*, lequel peut être conçu soit comme un agent intentionnel (« Mère Nature »), soit comme un processus mécanique, sans qu'il y ait de troisième option. Au contraire, le *teleological stance* de Gergely et Csibra permet de concevoir la sélection naturelle comme un agent téléologique non-intentionnel.

Nous avons vu que les physiciens adoptent l'attitude téléologique lorsqu'ils interprètent le comportement d'un système à partir de principes variationnels ou extrémaux, ce qui leur permet d'expliquer et de prédire son évolution en lui attribuant une fin, comme la minimisation du temps de parcours de la lumière ou la minimisation de l'énergie de surface d'une goutte d'eau, sans pour autant traiter le système comme un agent intentionnel (voir CHAP. IX, SECT. 4, p. 312). Darwin adopte la même attitude lorsqu'il dit par exemple que « *la sélection naturelle s'efforce constamment d'économiser toutes les parties de l'organisme* » (1872, Chapitre 5). Si l'on interprète le processus évolutif de cette façon, ou bien en termes de maximisation de la *fitness* (Birch, 2015; Fisher, 1930; Gardner, 2009), ou bien encore, plus généralement, en termes d'optimisation, alors les difficultés que pose la formulation anthropomorphique de Dennett sont levées. Cette interprétation permet en effet d'attribuer des fonctions aux parties des organismes, contrairement à l'interprétation mécanique, et elle rend mieux compte du processus de la sélection naturelle que l'attitude intentionnelle.

En revanche, il est sans doute plus facile de concevoir les êtres vivants comme des artefacts si l'on adopte l'attitude intentionnelle à propos de la sélection naturelle, car nous avons vu que les artefacts ont en effet une forte dimension intentionnelle et sociale qui nous pousse à penser, comme William Paley, qu'ils ont été faits par quelqu'un pour quelque chose.

Par conséquent, on peut considérer l'évolution par sélection naturelle comme un bricoleur fabriquant des artefacts et lui attribuer les mêmes états mentaux intentionnels que les créationnistes attribuent à Dieu et les Raéliens aux extraterrestres. Mais nous savons en même temps que l'horloger est aveugle et qu'il n'essaie même pas de fabriquer des horloges. Pour mieux expliquer le processus évolutif, nous devons plutôt le considérer comme un système téléologique non intentionnel, ce qui nous permet de saisir des *patterns* et de formuler des prédictions sans pour autant lui attribuer des états mentaux. Mais nous savons aussi que les phénomènes que l'on observe à l'échelle macro et de façon rétrospective sont le fruit d'une multitude innombrable de petits événements à l'échelle micro. On peut donc aussi considérer l'évolution comme la

résultante d'une série de mécanismes causaux élémentaires où aucun *pattern* n'est plus visible et où les fins et les fonctions sont exclus.

Si les êtres humains peuvent être considérés aussi bien comme des agents intentionnels que comme des artefacts, alors on devrait pouvoir leur attribuer une intentionnalité aussi bien originale que dérivée, selon l'interprétation que l'on adopte[189]. Pour d'autres organismes plus simples, il peut être plus judicieux de les considérer comme des agents téléologiques plutôt que comme des agents intentionnels. La plante qui s'oriente vers la fenêtre, par exemple, peut donc se voir attribuer des fins intrinsèques ou extrinsèques selon qu'on la considère comme un agent ou comme un artefact, mais on ne lui attribuera pas de croyances parce que cette interprétation est superflue et n'a aucune valeur prédictive.

En appliquant à la téléologie la justification que Dennett donne de l'intentionnalité, on peut affirmer que *les fins qu'on attribue aux organismes biologiques sont réelles et objectives* (ce qui apparemment fait de nous des réalistes), *bien qu'elles ne puissent être discernées que depuis le point de vue de celui qui adopte une certaine stratégie prédictive et que leur existence ne puisse être confirmée que par le succès de cette stratégie* (ce qui apparemment fait de nous des interprétationnistes).

Reste une question importante. Tous les organismes peuvent être interprétés aussi bien comme des agents téléologiques que comme des artefacts, mais la plupart des artefacts ne peuvent pas être interprétés comme des agents téléologiques. Vouloir prédire le comportement d'une chaise ou d'un ballon en lui attribuant des fins intrinsèques ou des états mentaux serait manifestement absurde et inutile. Qu'est-ce qui détermine le succès de la stratégie dans certains cas et pas dans d'autres ? Ce n'est pas l'appartenance au domaine du vivant, car des robots peuvent être considérés comme des agents (Terada et al., 2008, 2007). Ce n'est pas non plus une question de complexité interne, car un simple cercle sur un écran peut aussi être considéré comme un agent. Certains indices comportementaux sont importants, comme le mouvement autonome, mais pas suffisants, car on n'attribue pas de fins intrinsèques aux robots aspirateurs qui se déplacent tout seuls dans la maison.

L'une des réponses possibles est liée au fait que les attributions de fins sont généralement associées à des attributions de valeurs. D'après la psychologie cognitive, l'une des caractéristiques des agents est le fait d'avoir des intérêts qui leur sont propres, tandis que les artefacts ne répondent qu'à des intérêts externes. Si la plante oriente ses feuilles vers la fenêtre pour maximiser son exposition à la lumière, c'est parce la lumière est bonne pour la plante. Si le petit cercle saute par dessus le rectangle pour s'approcher du grand cercle, c'est sans doute parce que c'est bon pour lui, bien que nous ne sachions pas pourquoi. Si le cœur bat pour

189 Dennett semble parfois refuser d'attribuer une intentionnalité originale aux êtres humains.

faire circuler le sang, c'est parce que c'est bon pour lui et pour l'organisme auquel il appartient, car il contribue à la fois au bon fonctionnement de l'ensemble et à sa propre alimentation en oxygène et en nutriments. En revanche, si l'aspirateur autonome parcourt le sol de toute la maison, c'est parce qu'il a été conçu pour cela et parce que le fait d'avoir un sol propre est bon pour nous, bien que l'aspirateur lui-même n'en tire aucun bénéfice. Nous y reviendrons au CHAP. XIII.

L'association entre les fins et les valeurs nous amène à considérer une autre différence essentielle entre les êtres vivants et les artefacts, à savoir que l'organisation des machines manufacturées est presque entièrement tournée vers quelque chose d'extérieur à elles-mêmes, quelque chose qui a une valeur pour nous. Au contraire, l'organisation des êtres vivants est presque entièrement tournée vers l'organisme lui-même, c'est-à-dire notamment vers son auto-production, son auto-conservation et son auto-reproduction. Chez les artefacts, la forme est au service d'une finalité différente et indifférente à cette forme. Chez les organismes, au contraire, la finalité se confond avec la forme qui la rend possible ; la relation entre les deux est circulaire. Nous y reviendrons au CHAP. XIV.

6. Indétermination radicale des attributions téléologiques

L'argument de Dennett que nous avons suivi pour justifier la téléologie biologique pose un problème d'indétermination radicale des attributions de fins et de fonctions[190]. Dans la série numérique S_2, nous savons qu'il y a un ordre, mais nous ne savons pas lequel. Nous en avons proposé quatre interprétations différentes et incompatibles : S_3, S_4, S_5 et S_6. Chacune d'elles implique un certain nombre de prédictions dont le succès ou l'échec est déterminant pour évaluer leur correction. Mais rien n'interdit que deux stratégies prédictives soient équivalentes quant à leur succès tout en étant incompatibles quant à leur interprétation. Dans ce cas, si l'on ne peut pas recourir à un processus ou à une structure sous-jacente pour lever l'ambiguïté, selon l'argument de Dennett, alors il n'existe aucun moyen de trancher entre les deux interprétations rivales. Qui plus est, selon la Théorie algorithmique de l'information de Gregory Chaitin, on peut montrer qu'une interprétation (un algorithme) est meilleure qu'une autre (en termes de « compression » de l'information), bien que plusieurs d'entre elles puissent être équivalentes (avec un même « taux de compression »), mais il est impossible de savoir laquelle est la meilleure en termes absolus (c'est-à-dire quel est l'algorithme minimal).

190 Berent Enç (2002) aborde ce problème dans le cadre de la téléosémantique. Voir aussi le traitement minutieux du problème de la discrimination dans le cadre de la conception étiologique que réalise Philippe Huneman (2013c).

En ce qui nous concerne, cela signifie que nous devons abandonner l'idée qu'il existe une unique réponse vraie à la question de la fonction d'un trait biologique :

> « We cannot begin to make sense of functionnal attributions until we abandon the idea that there has to be one, determinate, *right* answer to the question : What is it for ? And […] there is no deeper fact that could settle that question […] » (Dennett, 1987, p. 319)

Cela n'a rien à voir avec le fait qu'un même trait puisse avoir à la fois plusieurs fonctions, ce qui est souvent le cas et ne pose pas de problème particulier. Ce qui est en question, c'est le fait que lorsque deux attributions sont en conflit, parce qu'elles répondent par exemple à des définitions différentes du concept de fonction, une seule d'entre elles puisse être vraie. Un cœur a-t-il la fonction de pomper le sang s'il en est incapable ? Les protubérances latérales des pseudo-lions sont-elles des pattes bien qu'elles n'aient pas d'histoire ? Les manchots ont-ils des ailes ou des nageoires ? Le panda a-t-il un pouce ? Les séquences d'ADN dit « poubelle » qui ne codent pas pour des protéines mais qui ont cependant une activité ou des effets ont-elles pour autant des fonctions ? Il n'y a peut-être pas de réponse définitive ni déterminée à ces questions dont les réponses correspondent parfois à des approches et des pratiques disciplinaires différentes au sein de la biologie. Et le problème que souligne Dennett est que Mère Nature elle-même pourrait ne pas le savoir :

> « Mother Nature doesn't commit herself explicitly and objectively to *any* functional attributions; all such attributions depend on the mind-set of the intentional stance, in which we assume optimality in order to interpret what we find. » (Dennett, 1987, p. 320)

Cela veut dire aussi qu'il n'existe peut-être pas une unique définition vraie du concept de fonction. Nous avons signalé à plusieurs reprises que les conditions de vérité, de validité ou de correction des définitions en lice n'étaient pas établies. Nous ne savons pas même à quoi les confronter pour en déterminer la valeur : certains disent que c'est au discours des biologistes, d'autres à leurs pratiques, d'autres encore à la nature, et tous recourent finalement à l'intuition. Si le choix d'une définition détermine une attribution de fonction et s'il y a indétermination radicale des attributions de fonctions, alors les définitions elles-mêmes sont indéterminées. Celle que nous avons nous-mêmes proposée au CHAP. VII ne prétend donc pas être vraie ni supérieure aux autres et elle admet sans problème une pluralité de formulations concurrentes, comme cela est indiqué clairement dans l'introduction. Sa vocation est de rassembler et de généraliser, autant que possible, les aspects positifs des autres définitions analysées tout en évitant leurs défauts conformément aux conclusions que nous avons tirées au CHAP. III.

L'indétermination des attributions de fonctions est extensible aux attributions de fins, puisque les premières, selon notre définition, sont essentiellement des contributions aux secondes dans le cadre d'une organisation. Les fins qu'on attribue aux êtres vivants et aux systèmes qui les composent sont aussi réelles que les croyances qu'on attribue aux êtres humains, bien que les unes et les autres soient relatives à une interprétation. Lorsque celle-ci est évidente, la vérité de l'attribution n'est pas en question. Lorsqu'elle est problématique, avec plusieurs options concurrentes et incompatibles, il n'y a pas forcément de réponse déterminée à la question : laquelle de ces attributions est vraie ?

Cette conclusion est similaire à celle que nous avons tirée au à propos de la vie et du vivant (voir notamment le CHAP. V, SECT. 5). Les deux choses sont liées. Si les êtres vivants sont essentiellement des agents téléologiques, c'est-à-dire des systèmes auxquels on peut attribuer des fins intrinsèques — contrairement aux fins extrinsèques des artefacts —, et si l'attribution des fins est indéterminée dans certains cas limites, c'est-à-dire principalement dans les cas où l'intuition est prise en défaut, comme les créatures laborantines de la biologie synthétique, alors l'appartenance à la catégorie des êtres vivants est elle-aussi indéterminée.

Chapitre XIII

Approche valorative

La définition étiologique de Larry Wright est satisfaite dès que se met en place une rétro-alimentation positive — l'équivalent d'une auto-sélection, selon ses propres termes. Or, celle-ci ne dépend pas de la nature du trait ni du caractère bénéfique ou délétère de ses conséquences, ce qui donne lieu à des contre-exemples comme celui du tuyau de gaz toxique. On peut pallier ce défaut en limitant la rétro-alimentation à la sélection naturelle, mais cette solution réduit la généralité de la définition et soulève d'autres problèmes, comme celui des premières générations et des doubles accidentels.

De plus, la sélection naturelle n'offre pas une explication suffisante de la fonctionnalité d'un trait. Elle permet d'expliquer comment la photosynthèse a pu apparaître toute seule, sans recourir à un créateur, mais n'explique pas pourquoi elle est apparue indépendamment des dizaines de fois. Pour répondre à cette question, il faudrait mentionner les bénéfices qu'elle apporte, c'est-à-dire sa valeur. Or, la sélection naturelle nous dit que les organismes porteurs de ce trait ont eu comparativement plus de succès reproductif. De là, on peut inférer qu'elle a dû leur être bénéfique, mais cette valeur supposée porte sur des générations passées ; elle ne nous dit pas si la photosynthèse est actuellement bénéfique pour les plantes vertes sur le balcon.

Le problème des valeurs est qu'elles sont censées ne pas exister dans la nature, de sorte que la science ne peut pas y avoir recours pour expliquer les phénomènes naturels. C'est pourquoi l'un des enjeux de la naturalisation des fonctions, pour les partisans de l'approche étiologique, n'est autre que de justifier le discours normatif et valoratif de la biologie en montrant qu'il s'appuie en réalité sur un mécanisme causal qui ne laisse aucune part à la subjectivité.

L'approche valorative cherche quant à elle à justifier le langage téléologique en montrant que (1) les attributions de fins et de fonctions impliquent des attributions de valeurs, et que (2) les attributions de valeurs sont justifiées. Elle peut prendre au moins deux formes. La

première consiste à dire que les valeurs n'existent pas dans la nature, qu'elles sont relatives à l'observateur, mais que les attributions de valeurs sont néanmoins justifiables, d'une façon ou d'une autre. La seconde affirme au contraire l'existence de valeurs naturelles, c'est-à-dire de valeurs propres et intrinsèques aux objets naturels et en particulier aux êtres vivants. Pour en arriver là, ces auteurs doivent faire le lien entre la définition de la vie ou du vivant et la définition des fonctions.

La question qui nous intéresse ici n'est pas tant l'existence de valeurs dans la nature que la façon dont cela affecte l'objectivité des attributions téléofonctionnelles. Il y a vraisemblablement un lien entre les fins et les valeurs, de sorte que si quelque chose peut être considéré comme une fin, alors cette chose a sans doute aussi une valeur. Cependant, comme nous l'avons montré aux CHAP. IX & X, les explications téléologiques n'ont pas toujours besoin de recourir, même implicitement, à des notions de valeur. On peut donc penser que la justification des explications téléologiques ne dépend pas nécessairement de l'existence de valeurs dans la nature ni de la légitimité de leur attribution.

Ce que nous allons essayer de montrer dans ce chapitre, c'est que cette justification est indépendante de la posture que l'on adopte à propos des valeurs. Nous commencerons par celle qui nous semble à la fois la plus faible et la plus « subjectiviste », puis continuerons *in crescendo* vers des postures de notre point de vue mieux fondées et plus « objectivistes », sans pour autant nous identifier avec elles.

1. De l'objectivité des faits fonctionnels

La discussion de l'approche mentaliste nous a conduits a justifier l'objectivité des attributions de fins et de fonctions depuis une posture à cheval entre le réalisme et l'interprétationnisme. Nous allons aborder ici la même question depuis une posture explicitement réaliste comme celle que défend John Searle (1995) afin d'en examiner les conséquences.

Admettons avec lui que certaines caractéristiques des objets matériels leur soient intrinsèques, comme la masse et la température, tandis que d'autres sont relatives à l'observateur, comme le poids et la chaleur. On dira que les premières sont ontologiquement objectives, car elles existent dans la nature, tandis que les secondes sont ontologiquement subjectives, car leur existence dépend des états mentaux des observateurs. Admettons aussi que les valeurs, comme le beau et le bien, soient relatives à l'observateur et ontologiquement subjectives. Admettons finalement qu'il y ait, entre autres, des faits bruts indépendants de l'observateur, comme « La Lune orbite autour de la Terre », et des faits mentaux comme « Je veux un verre d'eau ».

De là il suit que les jugements de valeur correspondent à des faits mentaux. Autrement dit: le fait que la survie et la reproduction soient bonnes pour l'individu, pour l'espèce, ou pour quoi que ce soit d'autre, n'est pas ontologiquement objectif, mais relatif à un observateur. Or, selon l'approche valorative, les attributions de fonctions sont elles-mêmes relatives à des attributions de valeurs : si la fonction des battements du cœur est de pomper le sang, c'est parce que l'on admet généralement que la vie et la survie sont une bonne chose et parce que le cœur y contribue de cette façon. Par conséquent, si nous pensions que la valeur la plus importante dans le monde était de glorifier Dieu en faisant des bruits de coups, alors la fonction du cœur serait de produire des bruits de coups, et le cœur le plus bruyant serait le meilleur (Searle, 1995, p. 15).

Depuis cette perspective, la fonction d'un organe, de même que celle d'un artefact, n'est jamais une caractéristique intrinsèque. Elle est toujours relative à un observateur. En d'autres termes, la fonction n'est pas dans la nature, mais dans le regard que l'on porte sur elle. Et cette distinction est d'autant plus justifiée ici que les attributions de fonctions portent toujours en dernière instance sur des faits bruts (Searle, 1995, p. 121). Ainsi, l'activité du cœur est un fait brut que l'on peut expliquer en termes de processus causaux ; tandis que la fonction qu'on lui attribue est un fait mental qui consiste à qualifier et à situer ces faits relativement à un système de valeurs que nous portons. C'est ce qui nous permet de décrire des processus causaux en termes succès et d'échec, de dysfonction, de malfonction, de maladie, etc.

Comment justifier l'objectivité des fonctions depuis cette perspective ? Tout d'abord, il faut distinguer entre deux types de fonctions. D'un côté, il y a celles que l'on assigne directement à un objet ou un processus. Par exemple, lorsque l'on utilise un caillou comme presse-papiers. Dans ce cas, c'est nous qui décidons de la fonction de la chose. D'un autre côté, il y a celles qui dépendent des valeurs et des fins que nous attribuons aux choses, mais qui ne dépendent pas directement de nous quant à leur détermination. Par exemple, si nous considérons que la vie est une bonne chose et que l'une des fins des organismes est de survivre, alors nous devons reconnaître que la fonction du cœur est de pomper le sang, car c'est la seule de ses conséquences qui soit nécessaire à la survie de l'organisme qui le porte. Dans ce cas, la fonction n'est pas assignée, mais découverte.

On peut en effet découvrir une fonction dans la nature, mais seulement après avoir assigné à la nature un ensemble de valeurs :

> « We do indeed "discover" functions in nature. But the *discovery* of a natural function can take place only within a set of prior *assignments* of value (including purposes, teleology, and other functions). Thus given that we already accept that for a species there is a value in survival and reproduction, and that for a species there is a value

> in continued existence, we can *discover* that the function of the heart is to pump blood, the function of the vestibular ocular reflex is to stabilize the retinal image, and so on. » (Searle, 1995, p. 15)

La même idée peut être formulée en termes de moyens et de fins. Après avoir assigné une fin à quelque chose, on ne peut pas choisir ce qui constitue ou pas un moyen de cette fin, mais seulement le découvrir. Car les moyens ne dépendent pas de nous. Ils dépendent des relations qui s'établissent entre la situation initiale, la situation finale visée et l'ensemble des contraintes (logiques, physiques et autres) qui limitent les chemins possibles entre les deux.

Ces fonctions-là sont donc épistémiquement objectives — bien que par ailleurs, depuis la perspective de Searle, elles soient ontologiquement subjectives. Cela signifie que nous pouvons *connaître* les faits fonctionnels, de même que nous connaissons les faits bruts, et que nous pouvons aussi nous tromper à leur propos. Lorsque nous attribuons une fonction à un trait biologique, cette attribution peut être vraie ou fausse, indépendamment de ce que nous en pensons ; sa vérité peut être établie objectivement.

Cela signifie aussi qu'il peut y avoir une science des faits fonctionnels. De fait, il y en a plusieurs. La médecine, la psychologie, la sociologie, la linguistique, la biologie, l'archéologie et d'autres reposent en partie sur la connaissance et la découverte de faits fonctionnels. En revanche, la physique et la chimie portent exclusivement sur des faits bruts. Par conséquent, toujours selon cette perspective, la biologie serait plus proche des sciences humaines que des sciences naturelles.

Il y a cependant une différence notable entre les objets des sciences humaines et ceux de la biologie. Si les premiers ont des valeurs, c'est parce que des agents intentionnels humains les leur ont données, et on peut en rendre compte comme on le fait pour les fonctions techniques. Les sciences humaines s'occupent alors d'étudier ces valeurs, mais elles ne les assignent pas, elles les découvrent ou les constatent. Les valeurs humaines sont donc ontologiquement subjectives, mais épistémiquement objectives. Par contre, les objets de la biologie posent problème. On ne peut pas faire référence à un dieu créateur ni tenir compte des valeurs et des fonctions qu'ils ont pour nous (alimentaires, esthétiques, etc.). S'ils ont des valeurs, comme la survie et la reproduction, elles doivent leur être propres. Autrement dit, c'est *pour eux* que la survie et la reproduction doivent avoir une valeur. Mais sur quoi faire reposer cette attribution de valeur si ce n'est sur une intentionnalité qui leur fait défaut ? Et comment pouvons-nous savoir que cela est bon pour eux ? Autrement dit, dans quelle mesure ces valeurs qu'on leur attribue sont-elles épistémiquement objectives ?

On peut interpréter la posture de Searle comme si les valeurs biologiques étaient attribuées par les biologistes eux-mêmes de façon plus ou moins arbitraire. Ceux-ci semblent en effet se mettre d'accord pour dire

que la survie et la reproduction sont bonnes pour les êtres vivants, et ils décrivent les traits biologiques en termes fonctionnels à la lumière de ces valeurs, mais ils pourraient tout aussi bien, semble-t-il, leur attribuer des valeurs différentes.

> « If we valued death and extinction above all, then we would say that a function of cancer is to speed death. The function of aging would be to hasten death, and the function of natural selection would be extinction. In all these functional assignments, no new intrinsic facts are involved. As far as nature is concerned intrinsically, there a no functional facts beyond causal facts. The further assignment of function is observer relative. » (Searle, 1995, p. 15-16)

> « We can, arbitrarily, define the "function" of biological processes relative to the survival of organisms, but the idea that any such assignment of function is a matter of the discovery of an intrinsic teleology in nature, and that functions are therefore intrinsic, is always subject to a variant of Moore's open-question argument: What is so functional about functions, so defined? Either "function" is defined in terms of causes, in which case there is nothing intrinsically functional about functions, they are just causes like any others. Or functions are defined in terms of the furtherances of a set of values that we hold—life, survival, reproduction, health—in which case they are observer relative. » (Searle, 1995, p. 16)

La lecture que nous faisons de ces passages est préoccupante et ne correspond peut-être pas à la pensée de l'auteur, mais si nous admettons que les valeurs n'existent que relativement à l'intentionnalité des agents, et que ni les êtres vivants ni la nature ne sont des agents intentionnels, et si nous admettons que les fonctions qu'on leur attribue sont relatives aux systèmes de valeurs que *nous* portons, alors qu'est-ce qui nous empêche de pouvoir en principe leur attribuer n'importe quelles valeurs, conformément à *nos* intérêts et à *nos* critères ? Si cette lecture est correcte, alors l'objectivité des attributions fonctionnelles en biologie, depuis cette perspective, s'en trouve considérablement affaiblie, pour ne pas dire complètement ruinée.

Cela pourrait nous conduire à penser que la biologie doit exclure de son discours les valeurs, les fonctions et la téléologie pour entreprendre finalement le sûr chemin de la science. Une autre option consisterait à accepter que les valeurs en biologie soient épistémiquement subjectives, mais qu'il existe certaines limites ou certaines contraintes qui font que l'on ne puisse pas attribuer arbitrairement n'importe quelles valeurs aux être vivants. Une troisième option consisterait à admettre que les êtres vivants eux-mêmes (et pas seulement les agents intentionnels) puissent être porteurs de valeurs, c'est-à-dire qu'il existe des biens naturels.

2. De l'objectivité des explications fonctionnelles

Quoi que l'on pense à propos de l'existence des biens naturels, on peut tenter de répondre au défi que pose la lecture du texte de Searle pour défendre l'objectivité des explications fonctionnelles dans le cadre d'une approche valorative. C'est ce que nous allons faire en nous inspirant des arguments développés par Andrew Woodfield (1976, p. 130-140) depuis une perspective différente de celle adoptée au chapitre précédent.

Tout d'abord, on peut montrer qu'il existe en effet des contraintes qui nous empêchent d'attribuer n'importe quelles valeurs aux êtres vivants. Il est évident, pour reprendre l'exemple de Searle, que la mort peut être considérée comme une valeur et une fin des organismes, mais il faudrait dans ce cas expliquer comment leur fonctionnement normal, c'est-à-dire l'activité normale de leurs parties, contribue à cette fin. Or, un organe comme le cœur contribue à la mort de l'organisme quand il cesse de battre, c'est-à-dire quand il ne réalise *pas* son activité normale. Il peut aussi y contribuer en réalisant son activité normale, à condition qu'il se passe quelque chose d'anormal ailleurs, comme une hémorragie. Pourtant, si la mort est une fin relativement à laquelle se définissent les fonctions des traits organiques, il faudrait que les activités normales de ces traits contribuent à cette fin, ce qui n'est pas le cas. Et pour que cela soit, il faudrait revoir toute l'organisation de l'organisme pour montrer que les activités habituelles de ses parties, qui contribuent *de fait* à sa survie, sont en réalité des activités anormales, dysfonctionnelles, tandis que celles qui ne se produisent qu'au bout de 40, 60, 80 ans ou plus, et en général une seule fois, sont quant à elles normales et fonctionnelles. Ce n'est tout simplement pas raisonnable.

De plus, les organismes ne sont pas des ensembles de parties indépendantes les unes des autres, mais des touts intégrés extrêmement complexes où les activités sont coordonnées et interdépendantes. Si nous pensions que la valeur la plus importante dans le monde était de glorifier Dieu en faisant des bruits de coups, et si nous pensions que c'est là la fin ultime des organismes, alors nous pourrions expliquer fonctionnellement les battements cardiaques depuis cette perspective, mais nous aurions alors beaucoup de mal à expliquer tout le reste. Et même si nous défendions que toutes les autres activités de l'organisme sont subordonnées et contribuent à rendre possible la production sonore du myocarde, la non-efficience de l'ensemble serait frappante. Pourquoi tant de complications pour si peu de bruit ? Par ailleurs, cette explication ne serait valable que pour les vertébrés. Comment rendre compte des fonctions (et de l'existence même) des traits de tous les autres animaux, plantes, champignons, bactéries, etc., qui n'ont pas de cœur battant ?

Revenons pour un instant à l'analyse fonctionnelle de Cummins qui consiste à expliquer les dispositions ou capacités d'un système en analysant de façon récursive le rôle des sous-systèmes qui le composent, c'est-à-dire en remontant à rebours les chaînes causales-dispositionnelles. On se souvient que cette analyse peut s'appliquer arbitrairement à n'importe quelle capacité de l'organisme, y compris à la production de bruits de coups, mais que son intérêt épistémique est proportionnel à la complexité relative et à la différence entre l'*explanans* et l'*explanandum*. Or le bruit des battements du cœur n'est ni différent ni moins « sophistiqué » que la production sonore de l'organisme, il en fait simplement partie. D'autres capacités sont beaucoup plus intéressantes à analyser fonctionnellement, car elles requièrent la coordination systémique de beaucoup d'éléments différents sur plusieurs niveaux d'organisation hiérarchique. Parmi elles, la survie et la reproduction sont sans doute celles qui font intervenir le plus grand nombre de traits organiques, à tous les niveaux, et qui impliquent la plus grande intégration systémique. C'est-à-dire que l'analyse de ces deux capacités-là permet de dresser le tableau le plus complet, le plus détaillé et le plus sophistiqué du fonctionnement de l'organisme dans son ensemble.

Tous les êtres vivants connus à ce jour semblent avoir en commun certaines capacités fondamentales, comme le métabolisme. Ce dernier rend possible la plupart des autres capacités de l'organisme ; il se situe donc au début ou presque de la chaîne des relations causales-dispositionnelles. À l'autre bout de la chaîne se situent la survie et la reproduction. Leur explication fonctionnelle permet donc non seulement de dresser un tableau très complet des interrelations causales-dispositionnelles au sein du système organique, mais elle permet aussi de le faire pour tous les êtres vivants connus, sans exception.

Puisque la survie et la reproduction se situent à la fin de la chaîne des relations causales-dispositionnelles, au sens où elles ne contribuent pas à une capacité de plus haut niveau du système, alors on peut les considérer *en ce sens* comme des fins du système, sans connotations mentalistes ni valoratives. Dès lors, l'analyse fonctionnelle d'un système peut avoir comme point de départ une fin, et pas seulement une capacité. Par exemple, la capacité de glorifier Dieu en produisant des bruits de coups serait aussi une fin *en ce sens* dans la mesure où elle se situe en bout de chaîne.

On peut réaliser une analyse fonctionnelle de l'organisme à partir de n'importe quelle fin, mais nous venons de voir que toutes les fins ne se valent pas. Certaines donnent lieu à des analyses plus complètes et plus intéressantes que d'autres en termes explicatifs ; elles produisent davantage de connaissances. Et depuis cette perspective, la survie et la reproduction jouissent d'une position épistémique privilégiée. Dès lors, même si le choix d'une fin était effectivement arbitraire, il n'en demeure-

rait pas moins que certains choix sont rationnellement meilleurs que d'autres : ils s'imposent à nous.

Cette conclusion est applicable au problème posé par la lecture du texte de Searle. Bien que l'on puisse en principe attribuer aux êtres vivants n'importe quelles valeurs, toutes les attributions ne se valent pas. La plupart n'ont aucune justification rationnelle et, parmi les autres, certaines sont manifestement meilleures que d'autres. Le problème est similaire à celui que pose l'attribution d'intentions aux êtres humains ou l'interprétation du sens des textes. En principe les possibilités sont infinies et nous ne saurons peut-être jamais laquelle est la correcte, mais il y en a souvent une ou deux qui s'imposent au dessus des autres.

Toutefois, si le choix des valeurs n'est pas totalement subjectif ni arbitraire, cela ne signifie pas pour autant qu'il soit objectif. Et le fait que le discours fonctionnel en biologie, en médecine et dans d'autres disciplines soit vecteur de connaissances ne justifie pas à lui seul l'usage scientifique de concepts semi-valoratifs en sciences ni l'idée qu'il existe des biens naturels. Après tout, l'hypothèse qu'un dieu a créé le monde est porteuse elle-aussi de connaissances véritables, comme les principes de Fermat et de Maupertuis dont nous avons vu que la justification initiale était explicitement théologique. Pourtant, cela ne suffit pas pour justifier la croyance en un dieu créateur, car les mêmes vérités auraient pu être découvertes sans cette hypothèse. Qu'une idée comme celle-là puisse être efficace n'implique pas qu'elle soit vraie. Il en va de même pour les valeurs naturelles, reconnaît Woodfield : « *One cannot justify a belief in natural goods by pointing to the heuristic or systematising advantage of holding such a belief* » (1976, p. 133).

D'après cet auteur, l'attribution de fonctions biologiques présuppose que les êtres vivants aient des biens naturels, sans préjuger de leur nature ni de leur existence objective. Si l'on regarde les fonctions comme étant objectives, alors on doit aussi regarder le bien d'un organisme comme étant objectif. De même, si l'on est subjectiviste à propos des biens naturels, alors on doit aussi l'être — du moins partiellement — à propos des fonctions. Est-il néanmoins possible, se demande-t-il, que les explications fonctionnelles soient objectives, quand bien même les attributions correspondantes ne le seraient pas ?

Pour Woodfield comme pour Wright, attribuer une fonction à un trait biologique, c'est avant tout donner une explication de sa présence ou de son activité dans l'organisme. Cette explication prend la forme suivante : « *X does A in S because A contributes (or level-generates) F and F is good for S* ». On constate ici que l'*explanans* est composé de deux éléments : un rapport causal et un jugement de valeur.

Le premier fait que l'explication soit véritative : dire que le cœur bat parce que ses battements contribuent à la circulation est vrai ; dire qu'il bat parce que ses battements font du bruit est faux. De plus, l'explication est objective au sens où elle exprime ce que Searle appelle un fait brut. Si

les battements cardiaques cessaient de contribuer à la circulation, le cœur cesserait de battre, car il ne recevrait plus l'oxygène et les nutriments nécessaires à cette activité. En revanche, si les battements cessaient de faire du bruit, alors il ne se passerait rien et le cœur continuerait de battre, toutes choses étant égales par ailleurs. Cette relation entre l'activité cardiaque et ses effets ne dépend pas de l'observateur.

L'élément valoratif sert quant à lui à indiquer que la chaine causale aboutit à une fin de l'organisme. Les effets de la présence d'un trait qui contribuent à en expliquer la présence n'ont pas toujours de fonction. Par conséquent, pour pouvoir attribuer une fonction, il faut identifier parmi les effets du trait ceux qui sont bons pour l'organisme. Autrement dit, parmi toutes les explications vraies de la présence ou de l'activité du trait, seules celles dont la chaine causale aboutit à une fin naturelle sont des explications fonctionnelles. Mais une fois que la fin pertinente a été identifiée, le rapport causal peut être formulé entièrement sans référence aucune à des valeurs ni à des fins.

D'après Woodfield, cela signifie que l'explication causale demeure objective, bien que l'analyse des descriptions téléologiques comporte un élément valoratif qui permet d'identifier l'explication pertinente dans une situation donnée :

> « The role of the possibly non-objective term "good for S" is to indicate that the causal chain goes through an end of the system. The class of ends is demarcated by an evaluative criterion, but once the relevant end has been *identified* in a given case, then the evaluative description of it can be replaced by a non-evaluative description. The causal connections between the events and processes themselves are always objective, even though the descriptions which identify them may contain evaluative terms. There is a sense, then, in which natural functional TDs are objectively explanatory, even though their analysis contains an evaluative term. » (Woodfield, 1976, p. 139)

Ainsi, même s'il n'existait pas de fins naturelles ni de biens intrinsèques aux êtres vivants, cela n'invaliderait pas les explications apportées par les descriptions téléologiques. Car ces explications disent par exemple que le cœur bat parce qu'il pompe le sang, et cela reste vrai indépendamment de la valeur que l'on accorde à la circulation sanguine. En reprenant les termes de Searle, on dira que même si les fonctions biologiques étaient ontologiquement subjectives, les explications correspondantes sont quant à elles épistémiquement objectives et elles font référence à des relations causales ontologiquement objectives.

De plus, nous avons vu précédemment que même s'il n'existait pas de fins naturelles ni de biens intrinsèques aux êtres vivants, on ne pourrait pas arbitrairement désigner n'importe quel effet comme étant une fin. Il serait rationnellement absurde de vouloir considérer la production

de bruits de coups comme une fin biologique, tandis que cela semble au contraire justifié dans le cas de la survie et de la reproduction. Par conséquent, même si les fins qu'on attribue aux êtres vivants étaient relatives à l'observateur, on ne pourrait manquer de reconnaître qu'il existe à leur propos un accord intersubjectif assez large et que celui-ci repose en grande partie sur des contraintes rationnelles.

3. Vertus et défauts de l'approche de Woodfield

D'après Woodfield, toutes les descriptions téléologiques partagent une structure abstraite commune analysée dans Fig. 27. Les fonctions biologiques y correspondent à la quatrième ligne, où X fait référence à un organe, une partie ou un mécanisme interne du système S ; A fait référence à une activité caractéristique de X ; et F fait référence à une activité de X qui est causée ou générée causalement par A. La relation $A \Rightarrow F$ signifie que « A contribue (normalement et de façon caractéristique) à F », A et F étant des types, et l'élément normatif (*F is good*) signifie que F est bon pour S (dans des circonstances normales), soit intrinsèquement, soit parce qu'il contribue de façon caractéristique à un bien ultérieur. Ainsi, quand on dit que le cœur (X) bât (A) pour faire circuler le sang (F), cela veut dire qu'il réalise cette activité parce que (*because*) ses battements

ANALYSANDUM		**ANALYSANS**				
Explanandum	***Explanans***	***Explanandum***	***Explanans***			
				Intensional	***Causal***	***Evaluative***
S does B	in order to do G	S does B	because	S believes	(B ⇒ G	& G is good).
S does B	in order to do F	S does B	because		B ⇒ F	& F is good.
X does/has A	in order to do G	X does/has A	because	S believes	(A ⇒ G	& G is good).
X does/has A	in order to do F	X does/has A	because		A ⇒ F	& F is good.

S : System ; X : part of a system ; B : Behaviour ; A : Activity ; G : Goal ; F : Function/end .

Figure 27: Tableau d'analyse de Woodfield. Toutes les descriptions téléologiques ont une structure abstraite commune qui repose sur la combinaison d'un élément causal et d'un élément valoratif. On peut les diviser en quatre types, selon que l'on parle du comportement d'un système (lignes 1 et 2) ou de l'activité d'une partie (lignes 3 et 4), et selon que l'on explique ce comportement ou cette activité en termes intentionnels (lignes 1 et 3) ou non-intentionnels (lignes 2 et 4). La première ligne correspond à la description du comportement d'un agent doté de désirs et de croyances (ou d'états internes analogues à des croyances). Les trois autres lignes sont fonctionnelles. La troisième correspond aux fonctions des artéfacts. La dernière, aux fonctions biologiques.

contribuent à faire circuler le sang ($A \Rightarrow F$) et que la circulation est une bonne chose (*F is good*) pour l'organisme (S).

Cette approche présente plusieurs avantages. Elle est proche de celle de Larry Wright, puisqu'une attribution fonctionnelle y est entendue comme une explication de la présence ou de l'activité d'un trait, mais le recours à la bonté des conséquences lui permet d'éviter des contre-exemples comme celui du tuyau de gaz toxique[191]. Et elle est formulée en termes de types, ce qui lui permet de rendre compte de la normativité du discours téléologique, mais elle ne recourt pas à l'histoire pour justifier leur existence.

De plus, elle est compatible avec une interprétation dispositionnelle au sens où $A \Rightarrow F$ n'est pas situé temporellement et ne fait pas référence à un objet ni à un événement particulier. Par exemple, quand on dit que les battements du cœur contribuent à la circulation sanguine, nous parlons du (des) cœur(s) en général et pas de celui d'un individu en particulier, et nous ne voulons pas dire que cela se soit déjà produit par le passé ni n'ait lieu actuellement ni ne doive se produire dans l'avenir. Il s'agit d'une affirmation générale

Un autre aspect intéressant de cette conception est la prise en compte de la relation partie–tout. Le trait biologique X auquel on attribue la fonction F n'est pas considéré en lui-même, mais en tant que partie d'un organisme S, car c'est relativement à l'organisme que la fonction est bonne. Autrement dit, c'est le tout qui détermine la fonction des parties.

Elle tient également compte de la relation moyens-fin, car l'une des principales découvertes de cette analyse selon son auteur est que les fonctions sont relatives à des fins. Dans le cas biologique, on peut considérer la survie et la reproduction comme des fins relativement auxquelles sont définies les fonctions des activités et des parties de l'organisme. Ainsi, on dira qu'une fonction du système circulatoire (X) est l'alimentation des cellules en oxygène et en nutriments (A), car cela contribue directement à la survie de l'organisme (F). Mais on pourra également dire que la fonction du cœur (X) est de pomper le sang (A), car cette activité contribue directement à la circulation sanguine (F), laquelle contribue à la survie. C'est-à-dire qu'une fonction, de même qu'un moyen, peut à son tour être considérée comme une fin lorsque l'on passe d'un niveau de description à un autre.

En ce qui concerne les défauts de cette approche, nous voulons tout d'abord souligner une double ambigüité dans la relation explicative. La première concerne l'inclusion de l'élément valoratif dans l'*explanans* : le fait qu'un comportement ou un effet soient bons fait-il partie de l'expli-

191 Woodfield (1998) qualifie sa propre posture de « conservationnisme sémantique » qu'il oppose au « révisionnisme naturaliste » de Larry Wright et de Ruth Millikan.

cation ? Autrement dit, si l'on devait placer des parenthèses pour circonscrire la portée du terme explicatif « parce que », l'élément valoratif y serait-il inclus ou pas. À la lecture du texte, il nous semble que la bonté ne fait pas partie de l'explication, mais nous allons voir dans la SECT. 5 que Bedau(1992b) penche pour une interprétation différente à partir de laquelle il développe sa propre conception.

L'autre ambigüité porte sur ce qui constitue une explication valide. Woodfiel en mentionne deux pour rendre compte des battements cardiaques. La première s'appuie sur la sélection naturelle, c'est-à-dire sur des causes distales. La seconde sur des causes proximales, à savoir que les battements d'un cœur particulier — n'importe lequel — font circuler le sang et contribuent dans des circonstances normales à la survie de l'organisme qui le porte, de sorte qu'ils contribuent aussi à leur propre existence. Le problème est de savoir si cette seconde explication, qu'il appelle « ontogénétique », est applicable à des traits moins déterminants pour la survie, c'est-à-dire à des cas où le lien causal entre les conséquences d'un trait et sa propre continuité immédiate ne soient pas aussi évidents à établir. Par exemple, il ne fait aucun doute que la rate a plusieurs fonctions importantes dans l'organisme puisqu'elle contribue notamment à la filtration du sang et au système immunitaire, mais un individu peut vivre de nombreuses années après son ablation. On peut aussi mentionner l'appendice chez l'humain, la queue chez le chien, etc. Ces traits contribuent d'une manière ou d'une autre à la survie de l'organisme porteur, et c'est la survie du système dans son ensemble qui assure leur propre continuité en tant que parties, mais contrairement au cas des battements cardiaques, on peut difficilement dire que leurs conséquences ($A \Rightarrow F$) soient causalement responsables de leur présence chez un individu quelconque.

La question est de savoir si une explication valide implique ici nécessairement que le trait ou l'activité fonctionnelle joue un rôle causal dans sa propre genèse ou continuité, ou bien si sa contribution à une fin du système est suffisante. Dans le deuxième cas, l'explication est similaire à celle que l'on peut donner à propos d'un item dans un artefact où ce qui nous intéresse ce sont les raisons de sa présence (*what for?*) plutôt que ses causes (*how come?*). Dans ce cas, il suffit de mentionner l'utilité de l'item pour en justifier la présence. Et l'analyse de Woodfield est que les descriptions téléologiques des objets naturels suivent le modèle des artefacts sans faire référence à des désirs et des croyances.

4. Pourquoi notre conception n'est pas valorative

L'une des différences principales entre notre formulation et celle de Woodfield est que nous n'analysons pas les fins en termes de valeurs. En effet, il nous semble que les fins ne sont pas toujours bonnes ni tenues pour telles. Ainsi, lorsque les physiciens expliquent que les rayons de lumière parcourent toujours le chemin de moindre temps, ou que les gouttes d'eau adoptent une forme quasi-sphérique pour réduire leur tension de surface, attribuent-ils une valeur à la minimisation du temps et de l'énergie ? Veulent-ils dire, ne serait-ce qu'implicitement et inconsciemment, que la consécution de ces fins est une bonne chose ? Il nous semble que non. Lorsque Aristote affirmait que les graves tombent pour rejoindre leur lieu naturel, attribuait-il une valeur à cette fin ? C'est possible, mais nous n'en sommes pas convaincus. Notre formulation se veut donc suffisamment générale pour laisser ouverte la possibilité que les fins ne soient pas nécessairement chargées de valeur.

De plus, quand bien même les fins auraient nécessairement une valeur, il nous semble que l'attribution et la détermination des valeurs est plus problématique que celle des fins. Nous nous sommes efforcés de montrer aux chapitres précédents que l'identification d'une fin n'implique pas nécessairement de jugement subjectif et qu'elle peut s'appuyer sur des outils mathématiques. Nous ne connaissons pas de méthode équivalente pour les valeurs.

Par ailleurs, la théorie de Kelemen en psychologie cognitive (CHAP. VIII) et la conception mentaliste de Woodfield sont assez proches dans certaines de leurs conclusions. D'après eux, les explications téléologiques seraient d'abord issues du domaine des actions intentionnelles et des artefacts, puis généralisées aux autres domaines. Il est possible, dit Woodfield, que les énoncés fonctionnels en biologie aient perdu, au cours de l'histoire, la référence implicite aux croyances ou desseins de Dieu qu'ils avaient initialement. Cependant, il est également possible que la téléologie n'ait pas une origine mentaliste ou intentionnelle, mais qu'elle constitue un système cognitif autonome, conformément à la théorie de l'attitude intentionnelle de Gergely et Csibra. Dans ce cas, l'analyse mentaliste de Woodfield serait moins pertinente qu'une définition comme la nôtre où les fins sont un concept primitif, de même que les valeurs le sont dans la sienne.

Quoi que soient les fins, elles permettent d'expliquer les moyens mis en œuvre pour leur réalisation, car la relation moyens–fin est un mode inné et fondamental d'appréhension du monde qui nous entoure. Elle repose sur une relation invariante entre le contexte (contraintes), les actions (moyens) et les résultats (fins). Les fonctions et les fins sont donc explicatives en ce qu'elles permettent d'identifier des invariants, contrairement aux valeurs.

5. Objectivité des valeurs et évolution darwinienne

Jusqu'ici, nous avons essayé de montrer que même si les valeurs étaient ontologiquement subjectives, le langage téléologique pourrait néanmoins être épistémiquement objectif, du moins dans une certaine mesure. Pour aller plus loin, il faudrait défendre l'objectivité ontologique des valeurs naturelles.

C'est ce que fait Mark Bedau lorsque, citant Aristote et Francisco Ayala, il dit aspirer à un dépassement du naturalisme étroit qui permettrait aux valeurs d'occuper une place dans le monde objectif:

> « A broader view of nature, perhaps roughly Aristotelian in outlook, could reckon objective standards of value as part of the natural order. According to this broader form of naturalism, which would contrast with supernaturalism and would reject the miraculous in nature, values would be real ineliminable natural properties, subject to broadly scientific investigation. Making sense of this broadly construed naturalism might enable the many attractions of a naturalistic treatment of biological teleology to be realized. » (Bedau, 1991, p. 655)

Les valeurs naturelles sont généralement attachées aux êtres vivants et il semble absurde de vouloir les appliquer aux rochers et à d'autres objets inorganiques, si ce n'est peut-être de façon métaphorique (Bedau, 1991, p. 655). Dès lors, pour défendre l'objectivité des valeurs dans le monde naturel, il faudrait d'abord justifier l'existence d'une distinction objective entre le vivant et le non-vivant, puis expliquer pourquoi les valeurs ne sont attribuables qu'aux seuls êtres vivants et pas aux autres objets naturels. Il faudrait donc vraisemblablement comprendre ce que sont la vie ou le vivant pour arriver à comprendre pourquoi des valeurs leurs sont attachées.

Avant d'examiner la manière dont cet auteur s'attaque à ces problèmes, arrêtons-nous brièvement sur sa conception de la téléologie. Nous disions plus haut que l'analyse de Woodfield présente une ambigüité quant à la portée de l'explication : la bonté de *F* en fait-elle partie ou pas ? Les deux réponses possibles à cette question sont le point de départ de l'analyse de Bedau (1992b) qui distingue trois degrés de téléologie :

> [G1] A Bs in order to C *iff* A Bs and A's Bing contributes to Cing and Cing is good for A.

[G1*] A Bs in order to C *iff* A Bs because A's Bing contributes to Cing and Cing is good.

[G2] A Bs in order to C iff [A Bs because A's Bing contributes to Cing] and Cing is good.

[G3] A Bs in order to C iff [A Bs because A's Bing contributes to Cing and Cing is good].

La proposition G1* correspond à la téléologie biologique de Woodfield. Ses deux interprétations donnent lieu aux degrés suivants. Au second degré (G2), les conséquences de l'activité B ne peuvent pas être accidentelles, mais le bénéfice qu'elles apportent oui. Autrement dit, on explique pourquoi A réalise l'activité B en mentionnant ses conséquences C, et il s'avère que celles-ci sont bonnes. Au troisième degré (G3), en revanche, le fait que les conséquences soient bonnes fait entièrement partie de l'explication.

Le second degré, qui est la conjonction d'une conception étiologique avec une condition valorative, rend compte de la téléologie biologique lorsque l'on fait intervenir la sélection naturelle. Le troisième correspond à une explication téléologique à part entière (*full-blooded*). Elle requiert la présence de mécanismes téliques capables de tenir compte des conséquences bénéfiques d'une action et leur donner une efficacité causale, c'est-à-dire notamment — mais pas nécessairement — des mécanismes mentaux. Bedau considère que le caractère problématique de la téléologie en biologie réside dans le fait qu'on n'a pas su distinguer entre le second et le troisième degrés.

L'argument de l'auteur repose sur la capacité à attribuer des valeurs ou bénéfices à quelqu'un ou quelque chose. Si A est un agent ou un organisme vivant, alors le bénéficiaire est A. Si A est un organe faisant partie d'un organisme, alors le bénéficiaire est l'organisme qui le contient. Si A est un artefact, alors le bénéficiaire est la personne qui l'utilise.

Le problème, c'est que nous ne sommes pas toujours capables de dire si une chose peut être considérée comme un agent (bénéficiaire) ou pas. C'est le cas des virus. Nonobstant, Bedau assure que son analyse de la téléologie en termes de valeurs reste valable, car les objets pour lesquels nous ne sommes pas sûrs que le discours téléologique soit applicable sont aussi ceux auxquels nous hésitons à attribuer des valeurs.

On peut s'interroger sur ce qui fait qu'un être vivant soit le genre de chose qui peut être un bénéficiaire (Bedau, 1992b, p. 794), mais répondre à cette question, reconnaît l'auteur, demanderait une théorie

des valeurs qu'il ne prétend pas développer. C'est par ailleurs quelque chose, dit-il, que tout le monde semble admettre (Bedau, 1992a). À défaut d'une théorie des valeurs, il propose donc une théorie de la vie.

S'appuyant sur les recherches en biologie synthétique, il défend l'idée qu'un système vivant cellulaire minimal doit intégrer trois fonctions critiques : il doit être capable de maintenir une *identité* tout au long de son existence ; il doit utiliser l'*énergie libre* de son environnement pour assurer son propre maintien, sa croissance et sa reproduction ; et il doit disposer d'un *support informationnel héritable* susceptible de modifications au cours de la reproduction. En d'autres termes, il doit avoir trois sous-systèmes chimiquement couplés : un *conteneur*, un *métabolisme* et des *gènes*, lesquels doivent être entendus en un sens purement fonctionnel, car leurs implémentations matérielles peuvent prendre de nombreuses formes.

Pourquoi ces trois fonctions-là et pas d'autres ? Parce que ce sont les conditions minimales requises pour que puisse se développer un processus d'évolution sans limites. En effet, l'essence de la vie selon Bedau et d'autres auteurs est la capacité d'*évolution illimitée* (Bedau, 2007) ou d'*adaptation souple* (Bedau, 1996), c'est-à-dire la capacité illimitée qu'on les systèmes vivants de produire des solutions nouvelles et variées pour répondre de façon appropriée aux variations imprévisibles qui affectent leur survie, leur reproduction et leur épanouissement.

Il ne s'agit pas là d'une simple opinion, ni d'un présupposé, ni du résultat d'une analyse conceptuelle, mais d'une tentative de formulation théorique au sens scientifique du terme. Car le but est le savoir ce que la vie *est*, indépendamment de ce que les gens *pensent* qu'elle est :

> « Je ne suis pas intéressé par la signification du mot "vie" ainsi que sa traduction dans d'autres langues. […] Je ne suis pas davantage intéressé par l'analyse des conceptions courantes de la vie. Ce sont souvent des *a priori* qui continuent à évoluer au fur et à mesure des découvertes. C'est la nature profonde de la vie qui m'intéresse. […] En d'autres termes, je voudrais comprendre la nature de la vie de manière suffisamment large pour y englober toutes les possibles formes de vie. Un telle théorie serait en effet faillible et sujette à révision. Mais je persiste à croire que l'élaboration et la défense de théories provisoires sont une manière constructive d'éliciter le progrès vers de meilleures théories encore. » (Bedau, 2007, p. 8-9)

Bien que Bedau n'explique pas pourquoi seuls les êtres vivants sont porteurs de valeurs, il défend l'idée que la vie et les valeurs existent réellement dans la nature et il propose un critère objectif de démarcation entre le vivant et l'inerte. C'est un grand pas vers la justification du langage téléologique en biologie. Si sa théorie était vraie et son approche correcte, il ne resterait plus qu'à se mettre d'accord sur ce que signifient les explica-

tions téléologiques, car le problème de leur légitimité scientifique serait pour ainsi dire réglé.

6. Approche valorative et théorie de la vie

Le problème est que les arguments qu'emploie Bedau pour défendre l'approche valorative de la téléologie ne sont pas toujours en consonnance avec sa théorie de la vie. Ainsi, lorsqu'il discute l'approche darwinienne (Bedau, 1991), il essaie de montrer que la sélection naturelle est insuffisante pour rendre compte de la téléologie et qu'il faut la compléter par l'introduction de valeurs[192]. Pour sa démonstration, il imagine un monde totalement dénué de vie où, cependant, une « population » de cristaux d'argile manifeste les quatre traits indispensables au travail de la sélection naturelle : reproduction, variation, hérédité et adaptativité. Puisque ce monde est dénué de vie, et puisque les cristaux ne sont pas des êtres vivants, aucune téléologie n'y est à l'oeuvre. Il se demande alors pourquoi la téléologie est applicable aux objets biologiques et pas aux cristaux d'argile alors que les deux sont soumis à la sélection naturelle. Sa réponse est que l'on peut attribuer des valeurs aux êtres vivants mais pas aux objets inertes. Pourtant, cinq ans plus tard, il reprend l'exemple des cristaux d'argile pour affirmer que, puisqu'ils satisfont le critère d'adaptation souple qui, selon lui, est une condition nécessaire du vivant, alors il faut les reconnaître comme vivants :

> « These counterintuitive cases do not refute the hypothesis that supple adaptation is the underlying explanatory factor that unifies the diverse phenomena of life. If this hypothesis is true, and if populations of viruses and clay crystallites, autocatalytic networks of chemicals, and even human intellectual and economic system exhibits supple adaptation, they all deserve to be thought of as "living" for they all depend on the same underlying process. » (Bedau, 1996)

Si les cristaux d'argile sont vivants et ont des valeurs, conformément à sa théorie de la vie, alors la téléologie doit aussi leur être applicable. Et si la téléologie leur est applicable, alors cet exemple cesse d'être un contre-exemple de l'approche darwinienne et sa justification de l'ap-

192 Le mécanisme de sélection naturelle n'agit pas sur les traits héréditaires en fonction de leur valeur : « In the biological and non-biological cases alike, natural selection promotes features that contribute merely to a creature's survival. Thus, although it is true that survival is good for living creatures, this is irrelevant to the natural selection. Natural selection is blind to the goodness that supervenes on a biological creature's survival. Since the goodness of survival does not itself play a role in natural selection, biological teleology never surpasses grade two teleology. » (Bedau & Packard, 1992, p. 801-802)

proche valorative s'en trouve considérablement affaiblie. Par ailleurs, si les cristaux d'argile sont vivants mais n'ont pas de valeurs (ni de téléologie), alors le lien entre les deux notions est rompu et nous n'avons plus de critère objectif pour attribuer des valeurs à certains objets et pas à d'autres. Finalement, si les cristaux ne sont pas vivants, alors le critère d'adaptation souple ne rend pas compte de la différence entre vivants et non vivants, ou alors cette différence ne justifie pas l'attribution différenciée de valeurs. Quoi qu'il en soit, les arguments et les conclusions des deux articles semblent difficilement conciliables.

Il y a deux autres objections à l'approche valorative que pourrait présenter ce que Bedau (1991, p. 665) appelle un « naturaliste étroit » et auxquelles il répond. D'un côté, le naturaliste pourrait soutenir que les cristaux d'argile, quoique non-vivants et dépourvus de valeurs, peuvent néanmoins être téléologiques, mais cette posture conduirait, répond-il, à un pantéléologisme radical peu plausible. D'un autre côté, le naturaliste pourrait rejeter la téléologie chez les cristaux en limitant celle-ci aux produits de la sélection naturelle chez les êtres vivants, mais alors il lui serait difficile de dire ce qu'est un être vivant depuis la même approche naturaliste « étroite » sans tomber dans un cercle vicieux.

Une troisième objection naturaliste que l'auteur ne mentionne pas consiste à dire qu'il n'y a de téléologie nulle part et qu'il n'est pas plus justifié d'attribuer des valeurs aux systèmes biologiques qu'aux cristaux d'argile : soit parce que les cristaux d'argile, bien que vivants, sont dénués de valeurs ; soit parce qu'il n'y a aucune différence essentielle entre le vivant et le non-vivant.

Supposons cependant que les cristaux soient vivants. Supposons également que la différence entre le vivant et le non-vivant implique une différence de valeurs. Supposons enfin, comme le font tous ceux qui travaillent dans le domaine des origines de la vie, que celle-ci soit une propriété émergente à partir d'un état hautement organisé de la matière. Il resterait à expliquer comment les êtres vivants acquièrent leurs valeurs.

Bedau traite indirectement cette question dans un article où il confronte son approche valorative à celle de la cybernétique (1992a). Il cherche à montrer qu'une approche de la téléologie en termes de mécanismes auto-régulateurs est insuffisante, car en brouillant la distinction entre systèmes biologiques et non-biologiques, elle ne permet pas de distinguer les vrais des faux « goal-directed systems ». Un pendule n'est pas un vrai système téléologique, car il tend vers un état d'équilibre, mais l'auteur imagine un pendule assez sophistiqué pour remplir les conditions de l'approche cybernétique et être considéré — à tort — comme dirigé vers un but. Ce qui permet de distinguer le vrai du faux, conclut-il, ce n'est pas le mécanisme causal, mais les valeurs.

Cet argument est généralisable à d'autres mécanismes au-delà de la cybernétique. On peut en effet partir de n'importe quel système physico-chimique qui n'est pas réellement dirigé vers un but, puis le compliquer

progressivement jusqu'à ce qu'il remplisse toutes les conditions exigées par une théorie causale de la téléologie et être considéré — à tort — comme dirigé vers un but. Ce qui distingue un système physicochimique complexe d'un système téléologique, d'après Bedau, ne se trouve pas dans les mécanismes causaux qui expliquent leur comportement, puisque ceux-ci peuvent être très similaires, mais dans le fait que les systèmes téléologiques — c'est-à-dire les êtres vivants — ont des valeurs et des intérêts que n'ont pas les systèmes physicochimiques non vivants.

Le problème est que cet argument, comme le précédent, se retourne contre son auteur. Si aucun mécanisme causal ne permet de rendre compte de la téléologie caractéristique du vivant, cela peut vouloir dire qu'aucun mécanisme causal ne permet de rendre compte de la différence entre le vivant et le non-vivant, ou qu'aucun mécanisme causal ne permet de rendre compte des valeurs qui caractérisent le vivant. Dans le premier cas, l'objectivité de la distinction entre le vivant et le non vivant serait remise en question. Dans le second, c'est l'objectivité des valeurs qui serait remise en question. Dans un cas comme dans l'autre, le critère d'adaptation souple se révèle insuffisant.

Au CHAP. V, nous avons recouru à un argument similaire. Si nous faisions évoluer un système physicochimique vers des niveaux de complexité croissante, nous pourrions peut-être arriver à un point où il remplirait les conditions que nous attribuons généralement aux êtres vivants, mais nous ne découvrirons pas un seuil à partir duquel il devient vivant. Les êtres vivants sont dirigés vers un but et leurs parties possèdent des fonctions, tandis que les systèmes physicochimiques, aussi complexes soient-ils, n'ont ni fonctions ni fins, car ces notions sont étrangères à la physique et à la chimie. Par conséquent, il n'y a pas de mécanisme causal permettant de distinguer le vivant de l'inerte, il n'y en a pas non plus pour distinguer les systèmes téléologiques de ceux qui ne le sont pas. La réciproque de Bedau est que si l'on ne peut pas distinguer les systèmes vraiment téléologiques sur la base d'un mécanisme causal, alors on ne peut pas non plus distinguer les êtres vivants.

Pourtant, dans un autre article, Bedau & Packard (1992) assurent être en mesure de dévoiler (et de mesurer) le mécanisme causal par lequel des entités non seulement prennent vie, mais acquièrent leur téléologie (et donc leurs valeurs). En partant de l'hypothèse que la vie est fondée sur l'adaptation souple, ils montrent comment, à partir d'une population d'organismes artificiels, une forme d'évolution peut apparaître et se maintenir. Ils expliquent ensuite comment une téléologie dérive de cette évolution. Reprenant les termes de Ayala, ils lient d'abord l'explication téléologique à la notion d'utilité, puis affirment qu'une mutation nouvelle peut donner lieu à un comportement bénéfique pour l'organisme. Cela dit, le comportement ne devient télique que si son utilité intervient causalement en faveur de son autoconservation au sein d'une

lignée d'organismes. Et c'est cette autoconservation que les auteurs mettent en évidence dans leur modèle :

> « Telic activity waves reflect genes that persist because they contribute to strategies of proven usefulness. They are not merely useful; they persist because they are useful. Thus, the behaviors in these well-tested strategies are teleological (goal-directed), not merely functional. » (Bedau & Packard, 1992, p. 27)

Le problème avec leur démonstration, extrêmement intéressante par ailleurs, c'est qu'elle pose initialement ce qu'elle prétend découvrir. En effet, les entités mises en jeu sont conçues dès le départ comme une population d'agents (vivants) au sein d'une biosphère. De ce fait, la téléologie est elle aussi présente dès le départ dans le modèle. La démonstration ne peut donc en aucun cas être concluante. Les vagues d'activité (*activity waves*) du modèle ne reflètent un comportement téléologique, comme le prétend Bedau, que si nous sommes disposés dès le départ à expliquer le comportement de ces entités (qu'ils appellent des « bugs ») en termes téléologiques. Si, au contraire, à partir du même modèle ou d'un modèle analogue, on considérait au départ les « bugs » comme des entités inertes (chimiques ou autres), il est vraisemblable que l'on n'interprèterait pas les vagues d'activité en termes téléologiques.

Mark Bedau est l'un des rares auteurs qui abordent aussi bien le problème de la téléologie en biologie que celui de la définition du vivant, et sa démarche a le mérite d'être aussi bien théorique qu'expérimentale, et aussi bien philosophique que scientifique. Cependant, elle illustre aussi certains des problèmes que nous avons identifiés dans ce travail.

Tout d'abord, il présuppose que la vie et la téléologie sont des propriétés des objets biologiques, c'est-à-dire qu'elles se trouvent dans la nature et pas dans le regard. C'est ce qui l'amène à réclamer un dépassement du naturalisme étroit qui permette d'intégrer les valeurs dans notre conception de la nature. Mais cela montre aussi que son naturalisme est guidé par des considérations ontologiques et aussi peut-être par une conception étroite de l'objectivité scientique.

Ensuite, les propositions de Woodfield et de Bedau partagent avec l'approche étiologique certains des problèmes que pose le recours à la sélection naturelle. En particulier, celui des premières générations et des doubles accidentels, à savoir qu'un trait n'a pas de fonction s'il n'est pas le produit d'une histoire sélective. Cela signifie que les membres et les organes des pseudo-lions n'ont pas de fonctions, bien qu'ils soient par ailleurs indiscernables de ceux produits par la sélection naturelle. Cela signifie également que si l'exploration spatiale nous permettait de découvrir sur une autre planète un objet analogue à une montre, nous pourrions attribuer des fonctions à ses parties, sans avoir besoin de connaître l'identité de son créateur ni la façon dont il a été créé. Par contre, si sur cette même planète nous découvriions un objet analogue à

un être humain, nous ne pourrions pas attribuer de fonctions à ses organes si les « petits hommes verts » n'étaient pas le produit de la sélection naturelle, mais d'un autre mécanisme causal jusqu'ici inconnu.

Une solution possible à ce problème consisterait à admettre plusieurs explications possibles de la présence d'un trait. Philip Kitcher défendait ainsi que la présence d'un trait biologique est explicable à partir des pressions sélectives auxquelles sont soumis les organismes qui le portent, mais que la sélection naturelle n'est pas le seul mécanisme causal impliqué dans la réponse à ces pressions. Woodfield mentionnait quant à lui deux explications possibles de l'activité cardiaque : une explication dite « phylogénétique » impliquant la sélection naturelle, et une explication dite « ontogénétique », en termes de causes proximales, où le fait de pomper contribue à la survie de l'organisme et, par conséquent, à la continuité de l'activité cardiaque en son sein. Dans un cas comme dans l'autre, l'explication mentionne un mécanisme par lequel les effets (bénéfiques) d'un trait biologique contribuent à sa propre présence. On peut aller plus loin en considérant que les organismes vivants sont des systèmes qui s'auto-(re)produisent continuellement, car ils renouvellent sans cesse leurs parties. C'est la voie que nous allons explorer dans la prochaine section à travers la proposition de Peter McLaughlin (2001) pour justifier la téléologie et les valeurs en biologie sans recourir à l'évolution darwinienne.

7. Objectivité des valeurs et auto-reproduction

Même si la vie se caractérisait par sa capacité d'évolution illimitée, on ne pourrait pas en dire autant des êtres vivants individuels. Ces derniers, suivant une idée anticipée notamment par Locke, Buffon et Kant, sont des systèmes organisés qui se caractérisent par la capacité de s'auto-reproduire continuellement, c'est-à-dire qu'ils se réparent et régénèrent leurs parties, croissent et se répliquent tout en demeurant identiques à eux-mêmes. Pour cela, ils extraient de leur environnement les matériaux qu'ils intègrent à leur structure, ainsi que l'énergie que requièrent ces opérations. Ce sont des systèmes dont l'activité est orientée vers leur auto-reproduction et dont l'identité repose sur cette même activité.

Il existe aujourd'hui des systèmes — comme nous les êtres humains — pour lesquels il semble justifié d'attribuer des intérêts, des fins et des même intentions propres. Avant l'apparition de la vie, de tels systèmes n'existaient pas. Si l'on veut naturaliser les valeurs, la téléologie et l'intentionnalité, il faudrait arriver à comprendre quand et comment de tels systèmes ont pu apparaître. Il en va de même pour les fonctions.

On peut imaginer un scénario où les interactions chimiques spontanées sur la Terre primitive ont conduit à une complexification croissante des molécules présentes dans le milieu et à des dynamiques et des réseaux de réactions de plus en plus stables, jusqu'à l'apparition de systèmes chimiques capables de se répliquer et de proliférer. Bien qu'ils ne fussent peut-être pas encore vivants, ces réplicateurs étaient peut-être déjà soumis à une proto-sélection naturelle. Par ailleurs, ces systèmes chimiques ont aussi acquis la capacité de se réparer et de se re-produire ou régénérer eux-mêmes. À un moment ou à un autre, certains de ces systèmes ont finit par avoir l'ensemble des propriétés qui caractérisent les organismes vivants. D'après McLaughlin, c'est à partir de ce moment-là que l'on peut commencer à leur attribuer une identité et des valeurs :

> « In any case, an entity that repairs and replicates itself is at least prima facie a candidate for having a self. […] The systems, insofar as they are actively engaged in providing for themselves, can be said to possess some kind of ridumentary interests, such as welfare, self-preservation, or self-propagation. Something that is good for (instrumental to) their self-reproduction may be said to be good for them in a sense that does not apply to other kinds of systems. » (McLaughlin, 2001, p. 183)

C'est aussi à partir de ce moment-là, ajoute cet auteur, que l'on peut commencer à formuler des explications fonctionnelles comme celles que réclament les partisans de l'approche étiologique. En effet, si un système se répare ou se régénère lui-même régulièrement, alors on peut dire que ses parties sont présentes (ont été régénérées) à l'instant T_0 parce qu'elles ont contribué (au moins en partie) à la régénération du système à T_{-1}. En d'autres termes, les parties d'un système sont causalement responsables (en partie) de leur re-production dans ce même système. Un organe comme le cœur est donc présent dans l'organisme à cause des effets de sa propre présence dans ce même organisme.

Cette conception est compatible avec la définition de Wright, car elle propose un mécanisme de rétro-alimentation ou d'auto-sélection grâce auquel les effets d'un item contribuent à sa propre présence. Et contrairement aux formulations darwiniennes postérieures, où la fonction d'un trait particulier dépend des effets qu'on eu d'autres traits du même type dans un passé indéfini, celle de McLaughlin permet d'expliquer la présence du trait à partir de ses propres effets :

> « When we say, with (causal) explanatory intent, that the function of X (for S) is Y, we mean at least:
>
> 1. *X* does/enables *Y* (in or for some *S*) ;
> 2. *Y* is good for some *S* ; and

3. by being good for some *S*, *Y* contributes to the (re)production of *X* (there is a feedback mechanism involving *Y*'s benefiting *S* that (re-)produces *X*). » (McLaughlin, 2001, p. 140)

> « The particular item x_i ascribed the function of doing (enabling) *Y actually* is a reproduction of *itself* and actually did (or enabled) something like *Y* in the past and by doing this actually contributed to (was part of the causal explanation of) its own reproduction. » (McLaughlin, 2001, p. 167)

L'un des avantages évidents d'une telle formulation par rapport à celle des définitions darwiniennes est qu'elle permet d'éviter des contre-exemples comme celui des premières générations et du double accidentel. Un autre avantage est qu'elle définit les fonctions d'un item à partir de ce que cet item fait dans le système contenant, c'est-à-dire à partir de son rôle causal, à l'instar des définitions systémiques.

Un autre avantage de cette proposition par rapport à celle de Bedau est qu'elle ne dépend pas du problème de la définition de la vie. On peut identifier l'une et l'autre avec les deux grandes approches théoriques que nous avons analysées dans la troisième partie de ce travail, à savoir l'approche évolutionniste et l'approche systémique, mais la posture de McLaughlin lui permet de ne pas s'engager dans ce débat : les entités dont on peut dire qu'elles ont un bien propre sont celles qui maintiennent activement leur identité à travers le temps, quel que soit par ailleurs leur statut vital et quelles que soient les définitions « correctes » de la vie et du vivant. Cela permet à sa définition des fonctions de ne pas rester cantonnée au domaine de la biologie et lui évite de tomber dans le problème épineux des limites du vivant.

L'avantage par rapport à d'autres conceptions systémiques, et en particulier par rapport aux conceptions cybernétiques, c'est qu'elle repose sur un mécanisme causal beaucoup plus spécifique et moins courant que les mécanismes de compensation ou d'auto-régulation. Le problème de ces derniers est qu'ils sont présents aussi bien chez les êtres vivants que dans des artefacts simples comme un thermostat et dans des systèmes naturels abiotiques comme les étoiles auxquels personne n'attribuerait de fonctions. L'auto-reproduction est un phénomène beaucoup plus rare dans la nature et nous commençons à peine à pouvoir le recréer artificiellement, notamment dans le domaine informatique, c'est-à-dire pour l'instant de façon virtuelle. Cela ne veut pas dire que nous ne puissions pas trouver des exemples contraires à l'intuition, à savoir des systèmes qui s'auto-reproduisent mais auxquels nous n'attribuerions pas de valeurs ni de fonctions.

Dans l'ensemble, la proposition de McLaughlin nous semble donc plus convaincante que celle de Bedau. Nonobstant, elle présente aussi un certain nombre de problèmes qui la rendent finalement insatisfaisante.

Le premier problème est l'attribution de fonction à des traits qui ne contribuent pas à la régénération de l'organisme, mais à sa reproduction au sens habituel. Certains sont même néfastes pour l'organisme puisqu'ils entrainent directement sa mort ou impliquent un effort immense qui diminue ses chances de survie. La réponse de McLaughlin sur ce point consiste à dire que ces traits sont fonctionnels, mais pas de la même façon, car leurs fonctions sont extrinsèques comme celles des artefacts. Dans ce cas, l'agent externe pour qui le trait est bénéfique n'est autre que sa progéniture.

Ce qui n'est pas clair, malgré les explications de l'auteur, c'est comment un trait peut avoir une fonction dans la mesure où il contribue au bien d'un agent externe non-intentionnel qui n'est pas encore né. À la limite, on pourrait admettre que les traits qui contribuent à l'auto-réplication d'un organisme soient bons pour lui lorsque ses descendants sont des copies exactes de lui-même et que le bénéficiaire n'est pas un objet matériel particulier, mais l'organisation ou la structure abstraite dont lui et ses enfants sont des implémentations. Mais comment peut-on attribuer des intérêts propres à des organismes fils qui ne sont pas encore nés et dont l'identité diffère de celle de leur parent ? Et comment le bien de ces organismes en puissance peut-il déterminer la fonction des traits d'un organisme réel pour qui ils impliquent un coût mais pas de bénéfice ? L'auteur répond à la première question en disant que les organismes fils sont eux aussi des systèmes auto-reproducteurs et que, à ce titre, ils ont un bien. La seconde reste en suspend.

McLaughlin consacre un chapitre entier à analyser ce que signifie « avoir un bien » (*Having a good*), et sa conclusion est que la possession de valeurs est liée, non pas au fait d'être vivant, mais au fait d'être un système auto-reproducteur capable de conserver son identité. D'après lui, on peut donc admettre que tous les êtres vivants aient un bien propre, mais c'est aussi le cas des sociétés humaines et, pourquoi pas, de futurs artefacts. Le problème est qu'il y a d'autres systèmes qui semblent remplir les conditions de l'auteur pour avoir un bien, et qui intuitivement n'en ont pas : les vésicules autopoïétiques de Luisi et Varela, la flamme d'une bougie… Des systèmes auto-reproducteurs relativement simples comme ceux-là peuvent être décrits en termes purement physicochimiques. Il n'est pas nécessaire ni utile de leur attribuer des valeurs ni de recourir au concept de fonction.

De plus, il est difficile de voir comment une telle conception peut rendre compte des dysfonctions et des malfonctions. Si un organe malade ou malformé est incapable de contribuer à l'auto-reproduction de l'organisme et à la sienne propre, dans quelle mesure peut-on dire qu'il a une fonction ?

Par ailleurs, cette conception est encore plus restictive que les conceptions darwiniennes, car certains traits auxquels sans doute personne n'hésiterait à attribuer des fonctions ne peuvent pas ici être

considérés comme tels. En effet, elle limite l'attributions de fonctions à l'explication de l'existence et des propriétés des parties d'un système auto-reproducteur qui, parce qu'elles contribuent à l'auto-reproduction de ce système, contribuent à leur propre reproduction par le système (2001, p. 209). Pourtant, les structures fonctionnelles ne sont pas nécessairement régénérées au cours de la vie de l'organisme. Par exemple, on a cru pendant longtemps que le cerveau ne fabriquait plus de nouveaux neurones après un certain âge. Nous savons que d'autres structures, comme le cristallin de l'œil, ne se régénèrent pas. Les capacités de régénération varient d'ailleurs énormément dans le domaine du vivant, et on peut facilement imaginer une espèce au cycle de vie très court qui ne répare pas ni ne régénère ses tissus à l'âge adulte et qui consacre toutes ses ressources à la reproduction.

Finalement, la description que fait McLaughlin des êtres vivants comme systèmes auto-reproducteurs nous semble beaucoup trop sommaire. L'auteur lui-même reconnaît que la question de l'identité et de l'individuation des organismes et de leurs traits a besoin d'être approfondie, ainsi que les implications de l'extension de la notion d'auto-reproduction à la propagation. Son argument aurait gagné à s'appuyer sur des réflexions déjà engagées dans la même direction comme celles déjà dévelopées par la biologie théorique, la théorie autopoïétique de Varela & Maturana (1974), la thermodynamique des systèmes hors-d'équilibre de Ilya Prigogine, ou les travaux de Robert Rosen, de Stuart Kauffman et d'autres auteurs sur l'auto-organisation en biologie.

L'approche organisationnelle des fonctions qui s'appuie sur tous ces travaux permet d'aller au-delà de McLaughlin dans la justification du lien entre l'auto-reproduction, la conservation de l'identité, la téléologie et la normativité. C'est à cette approche que nous allons consacrer le dernier chapitre.

La raison pour laquelle les propositions de Bedau et de McLaughlin sont discutées ici et pas dans les chapitres consacrés aux approches étiologique, systémique et mixtes est qu'elles revendiquent l'existence objective de valeurs naturelles et qu'elles définissent la fonction d'un trait relativement au bien propre du système auquel il appartient. L'une et l'autre disent s'inscrire dans un cadre de pensée naturaliste et admettent que les systèmes téléologiques, ceux dont les parties ont des fonctions, sont apparus de façon naturelle dans le monde à un moment ou à un autre de l'histoire des origines de la vie. Nonobstant, l'une et l'autre impliquent un coût métaphysique que beaucoup de naturalistes ne seraient pas prêts à payer.

Ce coût est un engagement quand à l'existence objective dans le monde d'entités ayant un bien intrinsèque, pouvant être bénéficiaires et stopper ainsi la régression fonctionnelle des moyens et des fins. Il consiste à accepter qu'il existe dans le monde des entités ayant des fins qui leur

sont propres bien qu'elles n'aient pas d'esprit pour se les représenter. Il s'agit d'accepter qu'il existe dans la nature des fins qui ne sont pas relatives à l'observateur et qui ne sont pas non plus seulement le terme ou l'état final vers lequel tend un processus causal, car elles ont une valeur — pas pour nous ni pour Dieu ni pour le Cosmos, mais pour ces entités elles-mêmes.

Nous partageons l'opinion de McLaughlin lorsqu'il affirme que le coût métaphysique des explications fonctionnelles est plus élevé que ce que le naturalisme contemporain avait prévu, mais que ces engagements métaphysiques ne sont pas incompatibles avec lui (2001, p. 212). La sélection naturelle n'est pas suffisante, comme nous avons pu le montrer dans ce travail, car même si elle permet d'expliquer comment les objets que nous reconnaissons actuellement comme vivants ont pu apparaître et se développer, elle ne rend pas compte des fonctions qu'on attribue à leurs parties. Certes, on pourrait éviter d'en payer le prix en donnant une définition stipulative des fonctions en termes d'adaptation, de valeur adaptative ou de rôle causal dans un processus donné. Cependant, il est difficile de voir ce que nous aurions à y gagner, car nous pourrions alors aussi bien bannir le terme de fonction en biologie pour en être débarrassés. Il suffirait pour cela, dit McLaughlin, que nous définissions la vie en termes de conditions minimales pour la sélection naturelle. Malheureusement, les discussions portant sur la définition de la vie montrent que certains biologistes ont tendance à formuler leurs définitions en termes fonctionnels.

CHAPITRE XIV

Approche organisationnelle

L'approche organisationnelle des fonctions est un prolongement récent des approches mixtes qui cherchent à unifier les conceptions étiologiques et systémiques (voir CHAP. III). D'un côté, elle vise à expliquer la présence du trait fonctionnel à partir de ses effets ; de l'autre, elle conçoit la fonction d'un trait comme étant son rôle causal *actuel* dans le système auquel il appartient. Elle se distingue des approches précédentes par la restriction des attributions fonctionnelles à un type de systèmes particuliers, à savoir ceux capables de s'auto-maintenir ou de s'auto-(re)produire. Or, cette capacité est considérée par certains comme l'une des conditions nécessaires du vivant (CHAP. V). L'approche organisationnelle permet donc de répondre, du moins en partie, à la question du lien entre les fonctions et le vivant.

On peut rassembler les différentes formulations de cette approche en deux versions principales. La première, défendue par Schlosser (1998) et McLaughlin (2001), met l'accent sur le fait que les conséquences d'un trait fonctionnel sont causalement responsables de sa propre présence dans un système auto-reproducteur. La seconde, défendue par Collier (2000), Christensen & Bickhard (2002), Mossio, Saborido & Moreno (2009) et Toepfer (2012) met l'accent sur l'organisation du système et la contribution des traits fonctionnels à son auto-maintien. Nous allons nous concentrer sur cette seconde formulation et, sauf mention du contraire, c'est à elle que nous ferons référence en parlant d'approche organisationnelle des fonctions (AOF). Pour une discussion plus générale, voir Garson (2016).

1. L'organisation du vivant

L'approche (AOF) développée par Mateo Mossio, Cristian Saborido & Álvaro Moreno (2009) vise à rendre compte des dimensions téléologique et normative du concept de fonction en biologie. On peut la situer dans la ligne des approches mixtes analysées au CHAP. III dans la mesure où elle parvient à concilier certains aspects des conceptions étiologiques et systémiques. D'un côté, elle conçoit les attributions fonctionnelles comme des explications de l'existence des traits, à la manière de Wright (1973). De l'autre, elle conçoit les fonctions comme des dispositions *actuelles* des traits qui contribuent à une capacité systémique, à la manière de Cummins (1975). Plus précisément, elle conçoit les fonctions comme une classe particulière de relations causales internes à l'organisation des systèmes biologiques qui, parce qu'elles contribuent à l'auto-maintien du système dans son ensemble, déterminent étiologiquement les conditions d'existence des traits qui les portent. En ce sens, les fonctions sont donc *à la fois* dispositionnelles et étiologiques (Mossio & Saborido, 2016).

Le noyau conceptuel de l'AOF est une conception de l'organisation du vivant entendu comme système auto-maintenu ou système autonome en un sens bien précis (Moreno & Mossio, 2015; Mossio, Montévil, & Longo, 2016). Elle s'appuie sur un cadre scientifique et conceptuel au carrefour de la biologie théorique, de la théorie des systèmes complexes et de la thermodynamique des systèmes loin de l'équilibre. Ses auteurs font plonger les racines de cette conception jusqu'à Immanuel Kant (1790), lorsqu'il décrit le vivant comme un être *organisé et s'organisant lui-même* où « toute partie, tout de même qu'elle n'existe que par toutes les autres, est aussi conçue comme existant pour les autres parties et pour le tout, c'est-à-dire en tant qu'instrument (organe) » (§ 65). Cet être, qui peut être appelé une « fin naturelle », se rapporte à lui-même réciproquement comme cause et comme effet ; il manifeste ainsi ce que l'on pourrait appeler une causalité circulaire. C'est essentiellement en ces termes, comme nous le verrons plus loin, que l'AOF justifie la téléologique dans le domaine biologique.

La perspective kantienne a trouvé une continuité dans l'approche organiciste au XIX^e^ siècle et dans la biologie des systèmes au XX^e^. Mais l'AOF plonge aussi ses racines dans les travaux de Bernard (1865) sur le « milieu interne » puis de Cannon (1929) sur l'« homéostasie », ainsi que dans d'autres domaines scientifiques. Dans les années 1930 et '40, des physiciens impliqués dans le développement de la théorie quantique, comme Schrödinger (1944), s'intéressèrent à la spécificité des phénomènes biologiques vis-à-vis de ceux de la physique et se penchèrent sur le problème de l'organisation du vivant. Par ailleurs, les travaux menés par Prigogine sur les structures dissipatives ouvrirent la voie en physique aux recherches sur l'auto-organisation et la thermodynamique des systèmes loin de l'équilibre. Parallèlement, le mouvement de la seconde cyberné-

tique allait déboucher dans les années '70 sur la théorie autopoïétique de Maturana et Varela (1973) tandis que, au même moment, des auteurs comme Waddington, Rosen, Piaget, Pattee et Ganti contribuaient au développement de l'approche organisationnelle en biologie théorique. C'est depuis cette pespective que l'AOF élabore sa propre conception du vivant en termes d'*autonomie* :

> « Many of these authors put strong emphasis on the idea that the constitutive organisation of biological systems realises a distinctive regime of causation, able not only of producing and maintaining the parts that contribute to the functioning of the system as an integrated, operational, and topologically distinct whole but also able to promote the conditions of its own existence through its interaction with the environment. This is essentially what we call in this book *biological autonomy*. » (Moreno & Mossio, 2015, p. xxvi)

La théorie autopoïétique décrit l'organisation minimale du vivant à partir du concept de *clôture opérationnelle* qui caractérise un système d'opérations dont les produits ou les résultats demeurent dans le système et contribuent à leur tour à la réalisation d'autres opérations. Ainsi, une cellule est un réseau opérationnellement clos de processus de production de composants à travers lequel elle se produit elle-même de l'intérieur et se constitue comme une unité individuelle dans l'espace physique (voir Fig. 9, p. 154). Cette description abstraite laisse volontairement de côté les aspects énergétiques et thermodynamiques auxquels sont soumis les systèmes matériels. D'autres auteurs, comme Stuart Kauffman, ont exploré parallèlement la voie de l'auto-organisation dans le cadre de la thermodynamique en étudiant des réseaux de réactions chimiques collectivement autocatalytiques qui se maintiennent loin de l'équilibre et qui constituent un ensemble dont les composants et les processus dépendent les uns des autres quant à leur production et à leur maintien, et qui contribuent collectivement à déterminer les conditions d'existence de l'ensemble lui-même. Moreno, Mossio et leurs collègues s'appuient sur les travaux de ces auteurs pour proposer une conception où la clôture organisationnelle qui caractérise les êtres vivants se situe, non pas au niveau des processus ou des réactions, mais au niveau des *contraintes* du système.

En termes simples, le flux d'énergie qui traverse un système doit être contraint d'une façon ou d'une autre pour qu'un travail puisse être produit. Par exemple, dans une machine à vapeur, ce sont les parois du cylindre et le piston qui, en contraignant l'expansion du gaz dans une direction et avec une pression données, transforment l'énergie thermique en travail. Dans un système physique auto-organisé, comme les cellules de convection de Bénard, c'est la structure macroscopique du système qui, en contraignant la dynamique des molécules environnantes, contribue à son propre maintien. Dans un tel système, les contraintes sont à la

fois une condition du renouvellement du travail et le produit du travail. Dans un système biologique, un grand nombre de structures différentes agissent comme contraintes les unes par rapport aux autres et contribuent collectivement au maintien de l'organisation dans son ensemble. D'après l'AOF, la clef de l'organisation du vivant se trouve dans la clôture de ces contraintes, c'est-à-dire dans leur dépendance mutuelle.

Nous n'allons pas entrer davantage dans les détails, car ceux-ci ne sont pas strictement nécessaires pour la discussion que nous allons mener sur la téléologie et les fonctions. Toutefois, il nous semble important de souligner l'une des conclusions de ces auteurs concernant la causalité, à savoir que la clôture crée un type d'organisation doté de propriétés émergentes et de pouvoirs causaux ontologiquement irréductibles aux « régimes de causalité » à l'œuvre dans d'autres systèmes naturels :

> « […] the organisation of biological systems can be shown to realize a distinctive causal regime, that we labelled closure. Closure translates into contemporary terms the original Kantian idea, according to which living systems can be conceived as natural purposes, in which each part exists with respect to the other parts, such as the whole is able to self-maintain. In order for closure to be a legitimate scientific concept—and not just an epistemic shorcut—philosophical arguments must be provided in favour of its emergent and irreductible character with respect to the causal regimes at work in other classes of natural systems. […W]e argued that closure is to be conceived as the mutual dependance among a set of constituents, each of them acting as a constraint. Constraints are configurations which, by virtue of the relations among their own constituents, possess emergent properties enabling them to exert distinctive causal powers on their surroundings, and specifically on thermodynamic processes and reactions. When a set of constraints realizes closure, the resulting organization constitutes a *higher-level emergent regime of causation*, possessing irreductible properties and causal powers. In particular, closed organizations are able to self-maintain as a whole (whereas none of the constitutive constraints can do it) which, in turn, enables them to generate biological *functions*. » (Mossio, Bich, & Moreno, 2013)

Nous sommes ici face à une forme de naturalisme où la teleologie et les fonctions sont conçues comme des propriétés émergentes d'un type d'organisation causale très particulier et spécifique aux systèmes biologiques. Il n'y a pas de causalité rétrograde ni descendante, mais un régime causal néanmoins différent de celui des autres systèmes naturels. De notre point de vue, cette approche présente un double avantage. D'un côté, elle évite la promiscuité des définitions systémiques antérieures. De l'autre, elle évite l'épiphénoménalisme des définitions étiologiques.

2. Téléologie et auto-détermination

L'AOF défend l'idée que l'organisation des systèmes biologiques est intrinsèquement téléologique, c'est-à-dire que leur activité est orientée vers une fin qui n'est pas imposée de l'extérieur, à l'instar des artefacts, mais auto-déterminée par le système lui-même. Elle se distingue d'autres approches par sa façon de légitimer scientifiquement l'attribution de fins à des objets naturels.

Pour un certain nombre de biologistes, les êtres vivants ont effectivement une fin qui se trouve encodée dans leur génome, de sorte que la finalité apparente des phénomènes biologiques peut être comprise à la lumière de mécanismes moléculaires contrôlés par un « programme » qui en détermine à chaque instant les étapes et dont le résultat « final » n'est autre que l'organisme lui-même (voir CHAP. II, SECT. 1). Dans cette conception, la téléologie ou téléonomie du vivant repose entièrement sur l'un des sous-systèmes de l'organisme, son programme génétique[193].

L'AOF propose une conception différente où la téléologie, entendue comme auto-détermination, repose sur l'organisation du système dans son ensemble. Ici, il convient de noter un certain flottement entre deux versions différentes de cette même conception.

La première affirme que la circularité causale d'un système auto-maintenu — sa clôture organisationnelle — suffit pour naturaliser la téléologie et la normativité (Moreno & Mossio, 2015; Mossio et al., 2009; Saborido, 2012; Saborido, Mossio, & Moreno, 2011). D'un côté, lorsqu'un système S contribue à maintenir certaines des conditions nécessaires de sa propre existence, alors à la question « Pourquoi S existe-t-il ? » on peut légitimement répondre « Parce qu'il fait Y ». De cette façon, on explique l'existence du système en faisant référence à ses effets (et pas à ses causes). D'un autre côté, disent les auteurs, l'activité d'un système auto-maintenu est importante pour lui dans la mesure où son existence même en dépend. Cette importance intrinsèque, ajoutent-ils, fournit un critère naturaliste pour déterminer les normes que le système est supposé suivre : « *the activity of the system becomes its own norm or, more precisely, its conditions of existence are the intrinsic and naturalised norms of its own activity* » (Moreno & Mossio, 2015, p. 71).

La seconde version affirme quant à elle que la causalité circulaire n'est pas suffisante pour fonder la téléologie et que celle-ci repose sur une forme spécifique de clôture qui opère au niveau des contraintes (Mossio & Bich, 2014). Le cycle de l'eau, par exemple, est une chaîne circulaire de transformations (processus) sous l'effet de contraintes *externes*, tandis que les systèmes biologiques ont des contraintes *internes* qui rendent possible leur activité et qui sont à la fois maintenues par cette activité, de

193 Pour une critique de l'approche génétique du point de vue de l'AOF, voir Mossio & Bich (2014).

sorte qu'elles dépendent mutuellement les unes des autres pour continuer à exister. Selon cette seconde version, les systèmes biologiques sont donc auto-déterminés au sens où ils sont auto-contraints.

Jusqu'ici, une explication était considérée comme téléologique lorsqu'elle faisait référence aux effets d'un phénomène plutôt qu'à ses causes. C'est en ce sens que la plupart des auteurs impliqués dans le débat emploient effectivement le terme. C'est aussi en ce sens que s'expriment les partisans de l'AOF quand ils affirment que la clôture organisationnelle permet de naturaliser la téléologie, car à la question « Pourquoi X existe-t-il dans telle classe de systèmes ? », il devient légitime de répondre « Parce qu'il fait Y » (Moreno & Mossio, 2015; Mossio et al., 2009; Saborido, 2012). Cela correspond très exactement à la première condition de l'approche étiologique de Larry Wright (1973): « X *is there because it does* Z ». Tout l'enjeu de la naturalisation étant alors de montrer qu'il existe une forme ou une autre de circularité causale, de sorte que les effets mentionnés ne soient pas seulement des événements futurs (ce qui impliquerait une causalité rétrograde), mais aussi des causes passées : circularité inter-générationnelle par le biais de la sélection naturelle pour l'approche étiologique ; boucles de rétroaction pour la cybernétique ; auto-maintien pour l'AOF.

Pourquoi alors, selon Mossio & Bich (2014), la téléologie intrinsèque ne peut-elle être fondée que sur la clôture des contraintes, et pas seulement sur la circularité causale ? Leur réponse passe par la distinction entre deux régimes de causalité différents. Dans un système comme le cycle hydrologique, les éléments qui le conforment (lacs, rivières, nuages, pluie…) forment une chaine circulaire de transformations *matérielles* où chaque élément génère le suivant, mais les conditions d'existence et les contraintes qui affectent le système ne sont pas générées par lui ; elles sont extérieures à la chaine de transformations. La dynamique de la rivière, par exemple, est déterminée en partie par la pente, la forme et la nature du terrain qu'elle traverse, mais ces contraintes ne sont pas à leur tour déterminées par la rivière ni par le cycle. Certes, la rivière creuse son lit, mais le cycle de l'eau continuerait d'exister même si elle ne le faisait pas. Le cycle peut donc agir sur ses contraintes externes, mais son existence ne dépend pas de cette action. En revanche, les systèmes biologiques ont des contraintes internes (membranes, enzymes…) qui dépendent les unes aux autres, et c'est à leur niveau que se situe la clôture. Ainsi, ils peuvent déterminer leurs propres conditions d'existence. C'est en ce sens, d'après ces auteurs, que l'on peut parler d'auto-détermination, laquelle permet de fonder la téléologie intrinsèque :

> « The idea of intrinsic teleology, we submit, does not merely point to the realisation of a circular relation between causes and effects but, rather, to the situation in which the activity of a system, by producing some effects, contributes to specifying the conditions under which the circular relation as such can occur. It is in this pre-

> cise sense that the connection between teleology and self-determination is to be understood. By merely obeying (or, at best, modulating) the external constraints, the dynamics of the cycles fail in specifying their causal regime in that they simply realise it. Accordingly, cycles do *not* self-determine. Therefore, they are not teleological regimes. » (Mossio & Bich, 2014)

L'une des différences principales entre les deux versions tient au types de systèmes dont on peut considérer qu'ils possèdent les dimensions téléologique et normative. Selon la première, cela inclut certains systèmes physiques et chimiques auto-maintenus comme la flamme d'une bougie. Pour la seconde, il s'agirait seulement des systèmes auto-contraints ; mais les auteurs reconnaissent qu'il existe un débat ouvert parmi les spécialistes quant à la question de savoir s'ils existent en dehors de la biologie. Si c'était le cas, ajoutent-ils, alors il faudrait admettre que la téléologie et la normativité ne sont pas l'apanage des systèmes biologiques.

3. La flamme d'une bougie est-elle un système téléologique ?

Une des caractéristiques des explications téléofonctionnelles est qu'elles font référence à ce que la chose est *censée faire* même lorsqu'elle ne le fait pas ou en est incapable. Elles possèdent en effet une dimension normative dont l'AOF prétend rendre compte en disant que l'activité qui permet au système de s'auto-maintenir est sa propre norme :

> « Nuestro modelo organizacional defiende que el auto-mantenimiento constituye el régimen causal relevante en el cual las dimensiones teleológica y normativa de las funciones pueden ser adecuadamente naturalizadas. […] la actividad de un sistema auto-mantenido es relevante normativamente porque su misma existencia depende de los efectos de esta actividad. Esta mutua dependencia entre la existencia y la actividad, la cual es una propiedad característica de los sistemas auto-mantenidos, es capaz de determinar las normas que el sistema y sus partes se supone que han de seguir siguiendo un criterio intrínseco, dado que no está impuesto por un observador externo de acuerdo a alguna razón extrínseca, y naturalizado, pues está relacionado a las características de la naturaleza del sistema y no se deduce de ningún principio metafísico o moral previo. Las *condiciones de existencia* del sistema son interpretadas aquí como las *normas* de su propia actividad. » (Saborido, 2012, Chapitre 5)

Puisque cette conception de la téléologie et de la normativité va au-delà du domaine biologique — selon la version 1 — et inclut certains systèmes physico-chimiques, elle est applicable, défendent ses auteurs, à la flamme d'une bougie.

> « what the flame does is relevant and makes a difference for itself, since its very existence depends on the specific effects of its activity. The conditions of existence of the flame are the norms of its own activity : the flame must behave in a specific way, otherwise it would disappear. » (Moreno & Mossio, 2015, p. 71)

La flamme exerce une série de contraintes sur son environnement : elle maintient la température locale au-dessus du seuil de combustion, elle liquéfie la cire de la bougie puis vaporise celle qui remonte par capillarité le long de la mèche, et elle induit un courant de convection qui apporte de l'oxygène et évacue les produits de la combustion. L'existence de la flamme, c'est-à-dire la combustion, s'explique donc en partie par les effets de sa propre activité. Ainsi, on peut dire par exemple que « la flamme existe *parce qu*'elle vaporise la cire ».

Peut-on dire pour autant que la flamme existe « *pour* vaporiser la cire » ? Il nous semble que non. Peut-on dire qu'elle « *doive* vaporiser la cire », car autrement elle disparaîtrait ? Tout dépend de l'interprétation de ce prétendu « devoir » attribué à la flamme.

La première interprétation est assez triviale : la présence de cire en phase vapeur est une condition nécessaire de la combustion. Autrement dit : il *faut* qu'elle vaporise la cire « pour » ne pas disparaître. On dit aussi que l'oxygène « doit » se lier à deux hydrogènes « pour » former une molécule d'eau, ou que la Lune « doit » passer devant le Soleil « pour » projeter son ombre sur la Terre. Le problème est que l'emploi des verbes *devoir* et *falloir* n'a ici aucune implication normative, contrairement à ce que semblent dire les auteurs.

La seconde interprétation prend au sérieux l'attribution d'un devoir au sens normatif, mais ce n'est sans doute pas celle des auteurs, car l'activité auto-maintenue de la flamme est un état de fait, et en déduire qu'elle *doive* réaliser cette activité serait commettre un sophisme naturaliste.

La troisième considère la flamme comme un agent dont les actions sont dirigées vers un but et sont soumises à l'obligation (auto-imposée) de maintenir son existence. Cette interprétation est justifiable à partir de la théorie de l'attitude téléologique, en vertu de laquelle n'importe quel objet (comme un simple cercle sur un écran) peut être considéré comme un agent poursuivant un but lorsque certaines conditions sont remplies. Nous ne savons pas si cette interprétation correspond à celle des auteurs, mais c'est la seule — à notre connaissance — qui rende compte de la valeur normative attribuée aux conditions d'existence de la flamme. Reste à savoir si la flamme peut effectivement être considérée comme un agent, c'est-à-dire si elle remplit les conditions nécessaires pour que l'attribution téléologique soit valable.

Tout d'abord, il est intéressant de comparer l'activité de la flamme avec celle du cœur. Ce dernier est un muscle dont les contractions requièrent un apport en nutriments et en oxygène, ainsi que l'évacuation des déchets résultants. La première est une combustion qui requiert un

apport en carburant et en oxygène, ainsi que l'évacuation des produits résultants. Le cœur contribue, par le biais de la circulation, à satisfaire la demande en nutriments et en oxygène nécessaires au maintien de l'organisme. La flamme contribue à satisfaire la demande en carburant et en oxygène nécessaires à son propre maintien. Dans les deux cas, si la demande n'est pas satisfaite, le système se désintègre.

Au premier abord, il semble que l'action de la flamme obéisse au principe d'action efficiente, car elle fond et vaporise la cire au rythme qu'il faut pour perdurer longtemps. Toutefois, face à des variations dans leurs conditions de fonctionnement, la flamme et le cœur ne se comportent pas de la même façon. L'organisme compense le manque d'oxygène (en haute montagne, par exemple) en accélérant le rythme respiratoire et le rythme cardiaque (ce qui entraîne localement une augmentation de la demande), en augmentant la pression artérielle, en réduisant l'irrigation de certaines zones, en réduisant le volume de plasma sanguin, en augmentant le volume de globules rouges et leur masse, en augmentant la myoglobine, les mitochondries, les enzymes aérobies, la concentration des capillaires, en modifiant la taille du ventricule droit, etc. De même, si au lieu de réduire l'oxygène disponible, on augmente la demande (effort physique), l'organisme y répond par une augmentation de l'apport : accélération des rythmes cardiaque et respiratoire, hypertrophie du ventricule gauche, etc.

Nous avons vu au CHAP. XI, SECT. 7, que l'explication fonctionnelle de l'activité du cœur est cohérente avec le principe d'action efficiente et avec la théorie de l'explication de Woodward. Lorsque les contraintes varient (oxygène disponible, effort physique), l'organisme réagit par des variations compensatoires permettant de maintenir une relation efficiente entre les moyens (rythme cardiaque, etc.) et les fins (alimentation des cellules en oxygène et nutriments, élimination des déchets). En d'autres termes, il existe une relation de dépendance contrefactuelle invariante entre l'apport et la demande.

La flamme de la bougie ne compense pas les variations, elle les subit. Lorsque l'on fait varier de façon analogue la « demande » au niveau des réactions de combustion (en augmentant le point de fusion de la cire, ou son indice limite d'oxygène), on ne remarque aucune augmentation correspondante au niveau de l'apport. Les réactions de combustion varient en effet de façon linéaire avec l'apport en oxygène : plus il y a plus d'oxygène disponible, plus la flamme est vive ; au dessous d'un certain seuil, elle s'éteint. Le comportement de la flamme n'est pas cohérent avec le principe d'action efficiente. L'équilibre entre l'apport et la demande n'est pas invariant sous des interventions contrefactuelles. On ne peut donc pas dire que la flamme brûle « *pour* vaporiser la cire », car ce n'est pas un système téléologique.

La version 2 de l'AOF nous semble avoir raison lorsqu'elle défend que la téléologie intrinsèque et la normativité ne peuvent pas être fondées seulement sur la circularité causale. Cela ne veut pas dire pour autant qu'elle ait raison de les faire reposer sur la clôture des contraintes.

4. La téléologie intrinsèque repose-t-elle sur la clôture des contraintes ?

Admettons que tous les systèmes auto-contraints soient intrinsèquement téléologiques et qu'ils aient pour fin dernière leur propre existence. Admettons aussi que tous les êtres vivants connus à ce jour et reconnus comme tels soient auto-contraints. De là on peut en conclure que tous les êtres vivants connus et reconnus comme tels sont téléologiques.

Cela ne veut pas dire pour autant que tous les systèmes auto-contraints soient nécessairement vivants ni biologiques. En effet, les auteurs de la version 2 reconnaissent que le débat reste ouvert quant à l'existence de systèmes auto-contraints abiotiques (Mossio & Bich, 2014). Il est donc possible, admettent-ils, que la téléologie intrinsèque ne soit pas limitée à la biologie.

Cela ne veut pas dire non plus que tous les êtres vivants soient nécessairement auto-contraints. En effet, on peut imaginer des formes de parasitisme et de symbiose où la clôture des contraintes ne se réaliserait pas au niveau de l'individu, mais à un niveau plus élevé. Les virus, par exemple, sont des parasites obligatoires qui n'ont pas de métabolisme propre et qui dépendent de celui de leur hôte. S'ils étaient vivants, ils ne seraient pas auto-contraints. Nonobstant, la théorie défendue par l'AOF ne serait pas pour autant remise en question, car la clôture des contraintes peut se produire à un autre niveau que celui de l'organisme.

Reste à savoir si tous les systèmes intrinsèquement téléologiques sont nécessairement auto-contraints. Là, nous touchons un point sensible.

Supposons que les virus ne soient pas vivants. Nous avons pourtant tendance à les décrire en termes téléologiques et fonctionnels. En effet, nous attribuons des fins à ce qu'ils font (survivre et se reproduire) et des fonctions à leurs parties (membrane, gènes). Si ces attributions sont acceptables et si les virus ne sont pas auto-contraints, alors la téléologie intrinsèque ne repose pas sur la clôture des contraintes.

Il y a d'autres objets auxquelles nous avons tendance à attribuer des fins sans qu'on puisse pour autant assurer qu'ils soient auto-contraints ni qu'ils soient vivants. Dans l'expérience de Gergely & Csibra (2003), les enfants (et les adultes) attribuent au comportement du petit cercle une fin sans connaître son organisation interne, c'est-à-dire sans savoir comment il est capable de réaliser ce comportement. Et il n'est pas sûr,

bien que l'expérience ne dise rien à ce sujet, que les enfants et les adultes voient le petit cercle comme un être vivant — ni comme un artefact.

On pourrait penser que cette expérience porte sur les conditions subjectives d'attribution de fins (i.e. relatives à l'observateur), tandis que l'AOF s'intéresse aux conditions objectives de la finalité (i.e. relatives à l'objet). Autrement dit, l'expérience hongroise s'interrogerait seulement sur ce qui nous porte à *croire* qu'un comportement est téléologique, tandis que l'AOF s'interroge sur ce qui fait qu'un système naturel *soit* téléologique.

Toutefois, cette lecture ne nous semble pas justifiée. Si l'on se tourne vers l'approche cybernétique, dont l'AOF est le prolongement, on voit une distinction claire entre la description phénoménologique de la téléologie et son explication scientifique. La téléologie y est entendue comme une propriété de certains systèmes dont le comportement est orienté vers un but. Cette tendance se manifeste dans des systèmes aussi divers qu'un missile auto-guidé capable de poursuivre une cible mouvante ou le système sanguin capable de maintenir constante sa concentration en eau. Il s'agit de phénomènes empiriquement observables que tout le monde reconnaît comme étant dirigés vers un but. À partir de cette base empirique, la cybernétique propose alors une caractérisation objective de ces phénomènes en termes par exemple de plasticité (équifinalité) et de persistance (résistance aux perturbations), puis une explication causale en termes de mécanismes de compensation et de rétro-alimentation. Par conséquent, elle oppose un ensemble de phénomènes empiriques à leur interprétation théorique, selon le schéma classique en épistémologie.

De manière analogue, on pourrait être tentés de dire que le travail des psychologues hongrois porte sur la description phénoménologique de la téléologie et sur ses conditions d'objectivation à partir du principe d'action efficiente, tandis que l'AOF propose une explication théorique en termes d'auto-contrainte ou de clôture organisationnelle.

Cette analogie ne nous paraît pas non plus satisfaisante, car les phénomènes étudiés par les uns et les autres sont différents. Chez Gergely & Csibra, les sujets attribuent une finalité aux déplacements d'un cercle sur un écran, c'est-à-dire à son comportement externe. Et cette fin n'est pas liée à leur survie ni à leur auto-maintien. L'AOF ne s'intéresse pour sa part qu'à l'auto-maintien des systèmes biologiques, c'est-à-dire à leur activité interne, laquelle dépend de leur organisation. Elle permet d'expliquer comment les systèmes biologiques arrivent à se tenir activement loin de l'équilibre thermodynamique, mais elle ne rend pas compte de leurs comportements. Elle ne s'applique donc qu'à l'une des nombreuses formes possibles de la téléologie.

Quand Aristote disait que les corps lourds tendent à rejoindre leur lieu naturel, il lui semblait que leur comportement était dirigé vers une fin. Quand nous voyons un enfant qui court derrière un ballon, un chien qui aboie, un toxicomane qui se drogue, un saumon qui remonte une

rivière, une bactérie qui remonte un gradient de sucre, un terroriste qui se fait exploser ou un alpiniste qui culmine le Mont Éverest, il nous semble également que ces comportements sont dirigés vers une fin qui n'est pas toujours nécessairement l'auto-maintien des systèmes qui les réalisent. Et ces fins sont intrinsèques dans la mesure où elles ne sont pas dictées ni déterminées de l'extérieur.

L'AOF a donc une portée très générale, parce qu'elle s'applique en principe à tous les systèmes biologiques ; mais elle entend la téléologie en un sens très restreint, car elle porte seulement sur le régime causal interne qui rend possible leur auto-maintien et ne considère qu'un seul type d'explication téléologique, celui concernant l'existence d'un item X en vertu de ses effets Y.

L'AOF contribue à répondre aux deux principaux problèmes que pose la téléologie en biologie. Le premier est celui de son existence. Quand on observe avec attention les êtres vivants, on ne peut qu'être émerveillés au premier abord par la complexité, la sophistication et la perfection de leur organisation, laquelle est manifestement au service d'une fin. Il semble en effet évident que les yeux sont faits pour voir et les ailes pour voler. C'est cela qui pousse les William Paley d'hier et d'aujourd'hui à croire que le vivant est le produit d'une intelligence surnaturelle, car ils ne concoivent aucune autre explication possible de leur *design*. Pourtant, cette explication existe et elle est scientifique. L'approche organisationnelle en biologie fait partie de la réponse, de même que la théorie de la sélection naturelle. L'une et l'autre nous aident à comprendre comment de telles « machines » peuvent exister naturellement.

L'autre problème de la téléologie est celui de sa justification. L'une des prétentions de l'AOF est de légitimer le discours téléofonctionnel en biologie en le naturalisant. Depuis cette perspective, leur proposition est assez convaincante. La clôture des contraintes est en effet un type particulier de circularité causale qui permet de dire, dans un cadre scientifique, que l'organisation des systèmes biologiques est orientée vers une fin et qui permet en outre de déterminer objectivement cette fin.

Notre approche est différente. Elle consiste à dire : puisque la causalité et la téléologie sont *de fait* deux modes de pensée que nous employons dans le domaine scientifique, intéressons-nous à ce qui les caractérise, ce qui les distingue et ce qui justifie l'emploi de l'un ou de l'autre selon le contexte.

Pensons à une partie de billard américain. Pour placer une bille donnée dans un trou, les joueurs cherchent le meilleur moyen d'y arriver en fonction de la disposition générale des billes sur la table. Dès lors, il suffit d'observer la manière dont un joueur s'apprête à frapper pour en déduire son objectif. Selon la théorie de l'attitude téléologique, ce raisonnement repose sur le principe d'action efficiente, qui est un principe d'optimalité mais aussi d'invariance. Il établit un rapport entre les moyens, les fins et les contraintes, de sorte que si l'un de ces paramètres

changeait (par exemple, la disposition des billes sur la table), alors les moyens (la manière de frapper la bille blanche) devraient changer eux-aussi pour atteindre de manière efficiente la même fin (mettre la bille noire dans un trou).

Le principe d'action efficiente ne s'applique pas seulement aux comportements externes, mais aussi à l'organisation interne des systèmes. Ainsi, en observant la disposition des pièces d'un jeu d'échecs, on peut en déduire les objectifs et les stratégies les plus probables de chacun des joueurs. De même, en observant la disposition des pièces d'une machine, on peut en déduire ce dont elle est capable et les conséquences de ces capacités, et, parmi elles, la finalité de la machine est sans doute celle pour laquelle son organisation constitue le moyen le plus efficient. Le mécanisme d'Anticythère original était tout aussi efficace qu'un bloc de marbre de même masse pour atteindre certaines fins comme retenir une pile de papiers, caler un meuble ou être lancé à la figure de quelqu'un, mais il était de toute évidence moins efficient pour ces fins là, car trop cher et trop complexe, entre autres raisons, parce qu'aucune d'elles ne rendait compte de la forme et de la disposition très particulières de chacune des pièces qui le composent. En revanche, il était peut être l'un des seuls objets dans le monde capables de représenter le mouvement des planètes et de la Lune, de prédire les éclipses et de déterminer le calendrier des Jeux Olympiques. Et cette capacité dépendait de l'organisation très précise de ses parties. Dès lors, la finalité qui rend le mieux compte de cette organisation et en vue de laquelle il est le plus efficient est sans doute celle d'un calculateur astronomique.

Le principe d'action efficiente est applicable à l'organisation des systèmes naturels comme à celle des artefacts. Si nous attribuons des fins et des fonctions aux êtres vivants et à leurs parties, ce n'est pas parce qu'ils sont auto-contraints, ni auto-maintenus, ni parce que leurs effets expliquent leur existence, ni même parce qu'ils sont complexes, mais tout simplement parce qu'ils satisfont le principe d'action efficiente. Et si ces attributions sont acceptables scientifiquement, ce n'est pas parce qu'elles sont traduisibles en termes de relations causales, mais parce que — comme nous avons essayé de le montrer aux CHAP. IX, X & XI — elles n'ont pas d'implications ontologiques inacceptables, elles ont une valeur explicative propre et complémentaire de celle des explications causales, car elles permettent notamment d'identifier des invariants différents de ceux des explications mécaniques, et elles contribuent de fait au progrès des connaissances en biologie et dans d'autres domaines, car elles rendent possibles des catégorisations, des inférences et des prédictions qui autrement ne le seraient pas.

5. Définition organisationnelle des fonctions

Jusqu'ici, nous nous sommes limités à discuter la façon dont l'AOF aborde la question de la téléologie, mais pas celle des fonctions. Or, d'après cette théorie, la fonctionnalité d'un système requiert, outre sa clôture, une différentiation organisationnelle, c'est-à-dire que ses parties doivent y jouer des rôles différents, complémentaires et hiérarchiques :

> « Functions, we submit, involve the fact that self-determination is achieved through the interplay of a network of mutually dependent entities, each of them making different yet complementary (and also hierarchical […]) contributions to the maintenance of the boundary conditions under which the whole system can exist. In other words, to ascribe functions we must distinguish between different causal roles in the system, a division of labour among the parts. And, of course, this is precisely what happens when closure of constraints is realised. » (Moreno & Mossio, 2015, p. 72)

La fonctionnalité d'un système est donc directement associée à son organisation, laquelle apparaît, selon les auteurs, lorsqu'il y a clôture des contraintes. Les trois concepts sont intimement liés :

> « "Functionality", "closure", and "organisation" are then *mutually related concepts*, which refer to the very same causal regime; in other words, in the autonomous perspective an organisation is by definition closed and functional. » (idem)

À partir de là, ils formulent une définition du concept de fonction selon laquelle : *un trait T a une fonction si, et seulement si, il exerce une contrainte sujette à clôture dans une organisation O d'un système donné* (Moreno & Mossio, 2015, p. 73; Mossio et al., 2009). Cette définition implique la satisfaction de trois conditions :

> A trait T has a function if and only if:
>
> C_1. T exerts a constraint that contributes to the maintenance of the organisation O;
> C_2. T is maintained under some constraints of O;
> C_3. O realises closure.

En ce qui concerne la première condition (C_1), il faut distinguer les contributions indispensables, comme celle du cœur, à défaut desquelles le système cesserait d'exister, et les contributions non indispensables mais qui affectent son organisation, comme les yeux, c'est-à-dire sans lesquelles le système peut continuer à exister en s'organisant autrement.

Suivant cette définition, on peut dire que le cœur a effectivement la fonction de pomper le sang puisque (C_1) cette activité contribue au maintien de l'organisme ; (C_2) le cœur est maintenu par l'organisme ; et (C_3) l'organisme réalise la clôture.

Les auteurs précisent que cette définition est très générale et vise à rendre compte de tous les cas possibles, de sorte que, en pratique, les attributions fonctionnelles devraient tenir compte de la complexité de l'organisation, c'est-à-dire de ce qu'ils appellent les différents ordres de clôture d'un système ou ses régimes d'auto-maintien.

Il faut donc distinguer deux types de fonctions, selon le régime d'auto-maintien auquel elles appartiennent, et qui sont définies comme suit (Mossio et al., 2009; Saborido, 2012) :

> La *fonction primaire* d'un trait T est la contribution de T à l'auto-maintien du système S faisant partie de la clôture organisationnelle du régime d'auto-maintien le plus basique de S.

> La *fonction secondaire* de T est la contribution à l'auto-maintien de S faisant partie de la clôture organisationnelle de n'importe quel autre régime d'auto-maintien (comparativement) plus complexe.

Ainsi, le cœur a une fonction primaire qui est de pomper le sang, et on peut éventuellement lui attribuer des fonctions secondaires, comme celle de produire un son caractéristique. Cet effet peut contribuer à un régime complexe d'auto-maintien (en aidant au diagnostic médical, par exemple) qui contribue indirectement à produire et à maintenir le cœur lui-même. Il convient de noter, précisent les auteurs, que les notions de fonction primaire et de fonction indispensable sont indépendantes. La fonction primaire du cœur est indispensable, mais celle des yeux ne l'est pas. Par conséquent, un même trait peut non seulement avoir une fonction primaire et une ou plusieurs secondaires, mais il peut aussi réaliser ces dernières et pas la première.

Nous n'allons pas entrer davantage dans les détails. Les objections principales que nous pouvons opposer à cette définition des fonctions dérivent de celles déjà exposées à propos de la téléologie dans la section précédente. Ajoutons à cela, en forçant un peu le trait, que le recours aux notions de « contrainte », de « clôture » et de « régimes d'auto-maintien » rent la définition difficilement applicable. Par exemple, quand on observe un trait sur le fossile d'une espèce disparue, comment savoir si elle exerçait « une contrainte sujette à clôture contribuant à un régime d'auto-maintien de complexité donnée de l'organisme et se trouvant elle-même maintenue par des contraintes exercées par lui » ? N'est-il pas plus simple de dire, comme Woodfield, McLaughlin et d'autres, que le trait existe parce qu'il contribue à la survie de l'organisme et donc, indirectement, à sa propre existence ? De plus, les auteurs de la théorie eux-mêmes ainsi

que les spécialistes de la question sont partagés quant à savoir si les structures dissipatives minimales comme la flamme d'une bougie sont des systèmes auto-contraints ou pas.

Par ailleurs, à titre de comparaison, la conception des fonctions que nous proposons se veut aussi générale que possible, tandis que l'AOF fait au contraire le choix de la restreindre au domaine biologique, bien qu'elle laisse en même temps la porte ouverte à une éventuelle généralisation à d'autres domaines (Mossio et al., 2009; Saborido, 2012). Tout au long de ce travail, nous n'avons cessé de répéter que les fonctions sont généralement associées au vivant et qu'il faut donc s'intéresser à ce qui distingue les êtres vivants des systèmes naturels non-vivants pour comprendre pourquoi les premiers sont porteurs de fonctions et pas les seconds. Or, c'est précisément ce que fait l'AOF en identifiant la clôture des contraintes et la différentiation organisationnelle. C'est-à-dire qu'elle situe la source de la téléologie et de la fonctionnalité dans l'organisation interne des systèmes biologiques. Le problème de cette solution, depuis notre perspective, est qu'elle ne permet pas de comprendre pourquoi nous attribuons aussi des fonctions aux comportements, aux artefacts, aux systèmes sociaux, etc.

Notre définition téléologique des fonctions (DTF) est compatible avec celle de l'AOF, mais plus générale et plus abstraite (Fig. 28). On constate que la condition C_1 de l'AOF implique une contribution de la fonction à un système, mais elle précise que cette fonction est une contrainte et que l'objet de sa contribution est le maintien du système, tandis que la DTF se contente de mentionner en C_2 une fin du système, n'importe laquelle. L'AOF précise ensuite que le trait doit être maintenu par le système (C_2) et que celui-ci doit être organisationnellement clos (C_3). Prises ensemble, les trois conditions de l'AOF impliquent que le trait, ou plutôt la contrainte exercée par lui, appartient au système défini par la clôture des contraintes qui le constituent. En effet, ce qui est caractéristique de l'organisation des systèmes biologiques, selon cette approche, ce n'est pas la clôture causale au niveau des processus et des structures (traits), mais la clôture au niveau des contraintes (fonctions). La DTF se contente de reprendre cette implication comme condition C_3 en disant que la fonction (et pas le trait) doit appartenir au système. De fait, cette condition d'appartenance s'inspire directement de la notion de clôture que l'on trouve dans les conceptions organisationnelles du vivant, et en particulier dans la définition autopoïétique, mais elle est beaucoup plus générale car on peut l'appliquer à d'autres systèmes dont l'identité n'implique pas la clôture. Finalement, la condition C_1 de la DTF s'inspire de celle de Wright et vise à éviter certains contre-exemples ; elle est implicitement présente dans la condition C_1 de la définition organisationnelle.

Définition téléologique	Définition organisationnelle
Le trait X a une fonction $f_{(x)}$ dans le système S et dans le contexte C si :	Le trait X a une fonction $f_{(x)}$ dans le système S si et seulement si :
C_1. $f_{(x)}$ est une conséquence de X ;	C_1. X exerce une contrainte $f_{(x)}$ qui contribue au maintien de l'organisation de S ;
C_2. $f_{(x)}$ contribue à une fin de S ;	C_2. X est maintenu sous certaines contraintes de l'organisation de S ;
C_3. $f_{(x)}$ appartient à S ;	C_3. S est organisationnellement clos (au niveau des contraintes).

Figure 28: Comparaison des définitions téléologique et organisationnelle. Leur formulation a été adaptée pour faciliter la comparaison. La définition organisationnelle peut être vue comme un cas particulier de la définition téléologique.

Conclusions de la cinquième partie

Selon l'approche mentaliste, les attributions téléologiques appliquées à la biologie et aux artefacts sont métaphoriques ou analogiques avec les actions humaines intentionnelles. Cela ne veut pas dire pour autant qu'elles soient fausses et superflues. On peut au contraire les prendre au sérieux et considérer qu'un système est véritablement dirigé vers un but s'il possède (objectivement) certains états internes que nous percevons (subjectivement) comme étant analogues à des désirs et des croyances. On peut également adopter une stratégie interprétative, à la manière de Dennett, en vertu de laquelle ce n'est pas un état interne du système qui détermine son caractère téléologique, mais le succès de la stratégie elle-même. Bien que relative à l'interprétation d'un observateur, la téléologie n'est pas arbitraire, car ce n'est pas lui mais la réalité qui décide du succès de son interprétation.

La théorie de Dennett inclut trois positions ou attitudes (*stances*) : physique, artefactuelle (*design*) et intentionnelle. Une théorie en psychologie cognitive défend l'existence d'une quatrième attitude, proprement téléologique, qui n'implique pas l'attribution de désirs ni de croyances. La combinaison des deux théories nous permet de justifier les attributions téléologiques sans dépendre des thèses mentalistes. Depuis cette perspective, on peut affirmer que les fins qu'on attribue en biologie sont réelles et objectives, bien qu'elles ne puissent être discernées que depuis le point de vue de celui qui adopte une stratégie interprétative et que leur existence ne puisse être déterminée que par le succès de cette stratégie. L'une des conséquences de notre posture est l'indétermination radicale des attributions téléologiques, c'est-à-dire qu'il n'y a pas toujours une réponse déterminée concernant la fonctionnalité d'un trait, ou la finalité d'un comportement, processus ou système.

Selon l'approche valorative, les fins impliquent des attributions de valeurs. Si ces dernières étaient arbitraires, les attributions de fonctions pourraient nonobstant être objectives — épistémiquement, mais pas ontologiquement — dans la mesure où la fonctionnalité d'un trait n'est pas choisie pas l'observateur, mais déterminée par sa relation à une fin. On peut cependant montrer que les attributions de valeurs ne sont pas totalement arbitraires, car certaines sont rationnellement meilleures que

d'autres. De plus, suivant Woodfield, on peut montrer que les valeurs servent à identifier parmi toutes les explications vraies de la présence d'un trait celles qui sont fonctionnelles, mais qu'elles n'interviennent pas dans les explications elles-mêmes, lesquelles peuvent en effet être formulées entièrement en termes de relations causales sans référence aucune à des valeurs ni à des fins. Cela signifie que les explications en question demeurent objectives, bien que l'analyse des descriptions téléologiques comporte un élément valoratif (non objectif) qui permet d'identifier l'explication pertinente dans une situation donnée.

Mark Bedau et Peter McLaughlin vont plus loin en essayant de justifier l'objectivité des valeurs dans le monde naturel. Le premier part du principe qu'il existe une distinciton naturelle entre les êtres vivants et les systèmes inertes, et que l'on peut attribuer des valeurs aux premiers, mais pas au seconds. Il ne justifie pas ce lien, qu'il tient pour évident, mais appuie sa distinction sur une théorie de la vie entendue comme processus d'évolution illimitée. Le second ne propose pas une théorie de la vie, mais tente de justifier l'existence objective de fins et de valeurs chez les systèmes capables de s'auto-(re)produire, comme les êtres vivants, sur la base d'une organisation orientée vers le maintien actif de leur identité.

Nous avons insisté dans la première partie de ce travail sur la place des valeurs dans les explications fonctionnelles. Pourtant, la conception que nous avons proposée dans la troisième partie n'est pas valorative et il y a plusieurs raisons à cela. D'abord, il nous semble que dans la plupart des cas — mais peut-être pas toujours — les fins sont chargées de valeur, de sorte que la référence à ces dernières serait redondante. Ensuite, bien qu'elles puissent nous aider à identifier les fins d'un système, les valeurs n'entrent pas dans la relation instrumentale entre les moyens et les fins qui caractérise la pensée téléologique. Par ailleurs, nous avons montré dans la quatrième partie que l'identification d'une fin n'implique pas nécessairement de jugement subjectif, et qu'elle peut même s'appuyer sur des outils mathématiques, mais nous ne connaissons pas de méthode équivalente pour les valeurs. Finalement, les fonctions et les fins sont explicatives en ce qu'elles permettent d'identifier des rapports invariants (contraintes–moyens–fins), contrairement aux valeurs.

Selon l'approche organisationnelle des fonctions (AOF), les êtres vivants sont des systèmes opérationnellement clos de processus ou de contraintes qui contribuent collectivement au maintien de l'organisation dans son ensemble. Les fonctions, la téléologie et la normativité sont des propriétés émergentes de l'organisation de ces systèmes dont le régime causal est différent de celui des autres systèmes naturels. L'AOF apporte une réponse convaincante aux deux principaux problèmes de la téléologie biologique : son explication (scientifique) et sa justification (naturaliste). Cependant, elle entend la téléologie en un sens très restreint qui n'est applicable qu'à l'une de ses formes possibles : l'organisation du vivant. Elle laisse de côté les comportements dirigés vers une fin et les artefacts.

CONCLUSIONS GÉNÉRALES (FRANÇAIS)

« The main finding is that functions are relative to ends. »

Andrew Woodfield (1976, p. 110)

Le concept de fonction est téléologique. Or, la téléologie en biologie est problématique. Pourtant, les biologistes continuent d'y avoir recours. L'un des enjeux du débat sur les fonctions, celui auquel nous avons consacré l'essentiel de ce travail, est alors de montrer que les attributions et les explications fonctionnelles en biologie sont néanmoins légitimes et acceptables.

L'une des stratégies, celle proposée par Robert Cummins, consiste à priver le concept de ses implications téléologiques. Les fonctions ne seraient, selon cette théorie dite du « rôle causal », que les dispositions causales des éléments d'un système qui contribuent à en expliquer les capacités globales. D'autres stratégies visent au contraire à justifier la téléologie biologique en la naturalisant. L'une d'elles, inspirée par la cybernétique, propose une explication scientifique du comportement des systèmes dirigés vers un but. Les fonctions pourraient dès lors être conçues comme des contributions à une fin, puisque la finalité d'un système peut à son tour être décrite en termes mécaniques. Une autre stratégie, proposée initialement par Larry Wright, consiste à interpréter les fonctions en termes de sélection. En biologie, cela signifie que les fonctions d'un trait sont ce pour quoi il a été sélectionné (ou a des chances de l'être), ou ce en quoi il contribue (au passé, au présent ou au futur) à la *fitness* des organismes qui le portent. Depuis cette perspective, attribuer une fonction à un trait serait une manière d'expliquer sa présence à partir du mécanisme de la sélection naturelle (ou d'autres mécanismes compatibles avec la théorie darwinienne). D'autres auteurs proposent de combiner plusieurs approches dans une conception unifiée.

Dans tous les cas, ces stratégies visent soit à éliminer, soit à réduire, soit à traduire ou encore à expliciter les attributions et les explications fonctionnelles sous forme d'attributions et d'explications causales. En pratique, malgré les différences quant à l'objectif poursuivi et aux moyens mis en œuvre pour y arriver, cela signifie que la plupart des auteurs proposent une définition du concept de fonction formulée en termes de relations causales.

Nous avons essayé de montrer que ces stratégies sont insatisfaisantes et inutiles, car elles proposent des solutions à un problème inexistant. En effet, les critiques lancées contre la téléologie pour justifier son exclusion des sciences de la nature sont infondées. Nous avons montré que les explications téléofonctionnelles n'ont pas forcément d'implications ontologiques ou métaphysiques inacceptables. La finalité n'est pas une forme invalide de la causalité, ni une projection abusive de la psychologie sur les objets naturels. Elle n'implique pas un engagement vis-à-vis de théories et de croyances contraires aux connaissances scientifiques actuelles, comme le vitalisme, l'animisme, le finalisme, le créationnisme, etc. De fait, les explications téléologiques sont employées en physique sous la forme de principes variationnels, sans aucune implication de ce type (puisque les mêmes phénomènes sont explicables en termes non téléologiques).

Nous avons montré également que les explications téléofonctionnelles sont acceptables du point de vue épistémologique. Elles sont dotées de force explicative et rendent possibles des catégorisations, des inférences et des prédictions qui, de fait, contribuent au progrès des connaissances biologiques. De plus, elles sont compatibles avec des théories modernes de l'explication comme celle de James Woodward que l'on peut généraliser pour qu'elle soit applicable aussi bien à la causalité qu'à la finalité.

La téléologie est un mode de raisonnement applicable à n'importe quel domaine d'objets qui permet d'identifier ce que Woodward appelle des relations de dépendance contrefactuelle invariantes. La finalité est un type de relation contrefactuelle invariante entre une action (les moyens), son résultat (la fin) et ses contraintes (la situation). Ainsi, par exemple, lorsque se produisent des variations dans les paramètres de la situation, les moyens doivent varier corrélativement pour pouvoir atteindre la même fin. En optique, cela signifie par exemple qu'un rayon de lumière adapte son trajet (les moyens) selon les indices de réfraction des milieux qu'il traverse (la situation) pour minimiser le temps de parcours (la fin). En biologie, cela signifie par exemple qu'un prédateur choisit sa proie (les moyens) selon sa valeur calorique et le temps ou l'effort nécessaires pour sa capture et sa consommation (la situation) afin de maximiser le rapport entre l'énergie apportée et dépensée (la fin).

En partant d'un cas simple (le mouvement d'un objet sur le plan) et en faisant varier l'environnement (présence d'obstacles), nous avons comparé les explications causales et téléologiques portant sur le même comportement. Nous avons montré que ce sont des explications diffé-

rentes et complémentaires : elles n'identifient pas les mêmes invariants, ne font pas les mêmes prédictions et n'ont pas la même généralité. Par ailleurs, elles sont toujours applicables toutes deux, mais dès que l'on complique un peu la situation, l'emploi préférentiel de l'une ou l'autre devient rapidement plus pertinent.

De façon générale, le raisonnement téléologique est fondé sur le principe d'action rationnelle ou efficiente. Ce principe consiste à attendre d'un agent qu'il accomplisse l'action la plus efficiente — parmi celles disponibles — pour atteindre ses fins, étant donné les contraintes de la situation. Autrement dit, lors d'une action dirigée vers une fin, la relation entre les moyens, les fins et les contraintes maximise l'efficience (laquelle peut se traduire par différentes variables : temps, argent, effort, énergie, etc.). Si ce n'est pas le cas, c'est-à-dire si l'action n'obéit pas au principe, alors l'interprétation téléologique n'est pas applicable. Par ailleurs, ce principe permet d'inférer l'un des éléments de la relation à partir de la connaissance des deux autres. Le même principe peut aussi être formulé par exemple en termes de problèmes–solutions–circonstances.

Ce n'est pas un principe ontologique ou métaphysique, mais seulement un outil cognitif à notre disposition pour interpréter le monde qui nous entoure, y détecter des régularités intéressantes et les exploiter à des fins pratiques (échapper à un prédateur, réparer le frigo, gagner aux échecs) ou théoriques (expliquer un comportement animal, prédire un phénomène physique). Selon certains chercheurs en psychologie cognitive, la pensée téléologique est un système de représentation inné indépendant de la pensée physique et peut-être antérieur à la pensée intentionnelle. Ce n'est donc pas, comme disait Piaget, une pensée causale immature. Ce n'est pas davantage une forme de mentalisme ou d'anthropomorphisme, car l'attribution d'une fin n'implique pas celle d'un esprit ni une forme ou une autre d'intentionnalité.

Il est vrai que beaucoup d'explications téléologiques sont fausses, infondées, voire totalement absurdes ou inacceptables d'un point de vue scientifique. C'est notamment le cas des explications religieuses des phénomènes naturels. Par exemple, certaines personnes pensent que le Soleil brille pour nous éclairer et que la terre tremble pour punir nos péchés. Il est toutefois facile de rejeter ces explications à l'aide du principe d'action efficiente, un minimum d'esprit critique et quelques données empiriques. Le problème ne vient pas de la pensée téléologique en elle-même, mais plutôt de ceux qui l'emploient, car ils commettent les mêmes erreurs grossières en ayant recours à des explications causales (pensée magique, superstition, parasciences, effet cigogne).

Si les explications téléologiques en biologie ne sont pas invalides ni inacceptables, alors le problème que les stratégies de naturalisation tentent de résoudre n'existe pas. Dans ces conditions, il est possible de définir le concept de fonction non seulement en termes de relations causales (causes–effets), mais aussi instrumentales (moyens–fins). Comme

disait Woodfield dans la citation en épigraphe, la principale découverte est que les fonctions sont relatives à des fins. Attribuer une fonction à un trait signifie que l'activité correspondante est un moyen d'une fin du système auquel elle appartient.

En essayant de formuler une définition téléologique du concept de fonction aussi générale que possible, nous avons abouti à la formulation suivante :

> La/une fonction d'un *token* de type X dans un système de type S et dans le contexte C est $f_{(X)}$ si :
> 1. $f_{(X)}$ est une conséquence de X ;
> 2. $f_{(X)}$ contribue à une fin du système $F_{(S)}$;
> 3. $f_{(X)}$ appartient à S.

Il nous semble que cette définition téléologique permet d'éviter la plupart des objections et des contre-exemples que l'on oppose habituellement aux autres. De plus, elle se veut suffisamment abstraite pour être compatible (autant que possible) avec un certain nombre d'entre elles entendues comme des cas particuliers d'un cas général. Par ailleurs, elle se veut applicable tant aux fonctions biologiques que techniques.

Une attribution fonctionnelle est à la fois une explication (non causale) de la présence d'un trait dans un système donné, à la manière de Wright, et une explication (causale) des capacités ou des fins de ce système, à la manière de Cummins.

Un aspect important de notre conception est la distinction entre l'architecture formelle d'un système (organisme ou artefact) et son architecture matérielle. La première correspond aux activités ou opérations qui sont collectivement constitutives du système (i.e. métabolisme, circulation, respiration, etc.). La seconde correspond à l'implémentation matérielle de ces opérations, c'est-à-dire aux structures concrètes qui les réalisent. Or, une même opération peut être réalisée par différentes structures. Ce ne sont donc pas les structures elles-mêmes qui ont une fonction, mais ce qu'elles font.

En attribuant une fonction à une opération, on peut expliquer les *raisons* de sa présence dans le système, indépendamment des *causes* de la présence de la structure matérielle qui la réalise. Par exemple, on peut expliquer rationnellement pourquoi le système circulatoire comporte une pompe, indépendamment du fait que celle-ci soit un cœur naturel, un cœur artificiel ou un circuit d'ECMO. Cela signifie aussi que, pour nous, contrairement à la conception darwinienne, les organes des doubles accidentels sont pleinement fonctionnels.

Par ailleurs, l'analyse fonctionnelle est ici entendue comme une rétro-ingénierie, laquelle permet d'examiner un système, comme le mécanisme d'Anticythère, dont on ne connaît pas forcément les capacités, ni l'utilité, ni l'histoire causale, pour essayer de comprendre simultanément comment il pouvait fonctionner et à quoi il pouvait servir étant donné

certaines contraintes, comme le contexte culturel de sa construction. Cela implique notamment que, contrairement à la conception de Cummins, les fins d'un système ne sont pas arbitrairement choisies par l'observateur. C'est l'*organisation* du système et son contexte (selon Kitcher : l'environnement, les pressions sélectives) qui permettent de déterminer ses fins et les fonctions de ses parties. Le critère d'optimalité joue un rôle important dans cette détermination.

Une autre manière d'aborder la question a été développée sous le nom d'approche organisationnelle des fonctions (AOF) par différents auteurs. De notre point de vue, la conception la plus intéressante est celle proposée par Mossio, Saborido, Moreno et leurs collègues. Elle s'appuie sur une théorie scientifique de l'organisation du vivant à partir de laquelle ils prétendent naturaliser la téléologie et la normativité biologiques. Ils montrent en effet, à l'instar de Kant, que l'activité du vivant est orientée vers une fin qui n'est pas imposée de l'extérieur, mais auto-déterminée par le système lui-même. Cette fin est l'auto-maintien ou l'autonomie. Les fonctions sont quant à elles une classe particulière de relations causales internes à l'organisation d'un système qui, parce qu'elles contribuent à l'auto-maintien de celui-ci dans son ensemble, déterminent étiologiquement les conditions d'existence des traits qui les portent.

Cette proposition nous rappelle celle de Rosenblueth, Wiener & Bigelow dans leur article précurseur de la cybernétique. Les uns et les autres tentent en effet de donner une expression scientifique rigoureuse et novatrice au concept de finalité. S'il fallait traduire les fins et les fonctions *biologiques* en termes de relations causales, il nous semble que l'AOF serait sans doute la meilleure option. Cependant, nous défendons ici que la téléologie n'a pas besoin d'être ainsi traduite pour être justifiée et que l'AOF ne rend compte, au mieux, que de l'une des formes de la téléologie, celle du vivant.

La fonction d'un trait est-elle déterminée indépendamment de nous ? Est-ce une propriété naturelle, que nous pouvons découvrir dans le monde (et à propos de laquelle nous pouvons nous tromper), ou une classe humaine relative à nos intérêts épistémiques et nos stratégies interprétatives ? Différentes conceptions des fonctions peuvent ne pas tomber d'accord sur le fait qu'un trait possède ou pas une fonction, et, au sein d'une même conception, l'attribution d'une fonction précise à un trait n'est pas forcément évidente. Depuis une posture réaliste, comme celle qu'adoptent les partisans de l'approche étiologique et aussi certains partisans de l'approche systémique, alors la réponse à la question de la fonctionnalité d'un trait se trouve dans la nature.

Étant donné que les fonctions et la vie sont étroitement liés, au sens où les êtres vivants (ou les objets de la biologie, au sens large) sont justement le genre de choses qui sont censées avoir des fonctions, tandis que les autres systèmes naturels (les cailloux, les nuages, les étoiles) n'en ont pas, nous nous sommes donc penchés sur les recherches portant sur la vie

minimale (biologie synthétique) et sur les origines de la vie, pour essayer de comprendre comment se fait l'acquisition des fonctions lors de la transition de l'inerte au vivant, et ce que sont au juste ces prétendues propriétés naturelles que l'on attribue à certains objets et pas à d'autres. Cela implique tout d'abord de pouvoir distinguer entre les uns et les autres. Parmi les questions que se posent les chercheurs, se trouvent celle des critères permettant de reconnaître une forme de vie nouvelle, que ce soit au fond d'une éprouvette ou sur une planète distante, et celle de la définition universelle de la vie et du vivant. Or, on constate que les stratégies adoptées pour tenter de répondre à ces questions sont similaires à celles du débat sur les fonctions (systémiques, darwiniennes et mixtes) et que les problèmes qu'elles rencontrent le sont aussi, avec notamment la difficulté à établir une distinction non-ambigüe entre la vie et la non-vie en employant des critères de complexité ou de sélection naturelle. De plus, les définitions qu'elles proposent sont souvent fonctionnelles, de sorte que leur application dépend de l'interprétation du concept de fonction.

Il n'y a peut-être pas de distinction naturelle entre la vie et la non-vie. Autrement dit, cette distinction n'est peut-être pas dans la nature, mais dans le regard que nous portons sur elle, c'est-à-dire dans nos catégories, nos modes de représentation et nos méthodes d'investigation. La transition de l'inerte au vivant ne se trouve peut-être pas dans la nature, mais dans la manière de décrire les systèmes naturels à mesure qu'ils deviennent plus complexes et acquièrent certaines propriétés pour lesquelles les outils de la biologie deviennent pertinents.

Si la vie et le vivant sont des catégories humaines, cela ne veut pas dire pour autant qu'elles soient arbitraires ni que la scientificité de la biologie soit remise en question. Les planètes sont aussi une catégorie humaine à laquelle les astronomes ont donné en 2006 une définition stipulative et conventionnelle dont la portée est limitée au Système Solaire, car les planètes extrasolaires ne sont pas définies. Avec elle, les astronomes ne prétendent pas dire ce que *sont* les planètes, mais décident d'employer ce terme pour désigner une classe de corps dont les propriétés sont intéressantes. Les planètes sont donc relatives aux intérêts épistémiques des astronomes. Elles ne sont pas pour autant arbitraires et l'astronomie n'en est pas moins une science rigoureuse.

Si la vie n'est pas une propriété naturelle, elle correspond peut-être à ce que Dennett nomme les propriétés « charmantes ». Bien que relatives à l'observateur, elles impliquent une forme de connaissance du monde extérieur, car elles sont l'expression d'une relation entre nos structures cognitives et les structures du monde perçu, lesquelles sont souvent la manifestation de structures plus profondes. Cette relation rend possibles les généralisations inductives, les prédictions réussies et les explications scientifiques. Il y a là une connaissance objective qui ne requiert pas un engagement ontologique vis-à-vis de la chose connue.

Cela nous amène à penser que les fonctions ne sont peut-être pas non plus des propriétés naturelles au sens où l'entendent Millikan, Longy ou Davies. Au contraire, nous défendons l'idée que les fonctions sont un type de relation propre de l'attitude téléologique, de même que la causalité est un type de relation propre de l'attitude physique. Les deux attitudes sont des modes innés, autonomes et complémentaires de représentation du réel qui peuvent *de jure* et *de facto* être employées dans un cadre scientifique. Bien qu'elles n'existent pas indépendamment de nos intérêts épistémiques et de nos stratégies explicatives, les causes et les fonctions peuvent néanmoins être objectives, car le succès de ces stratégies ne dépend pas de nous. Autrement dit, lorsque nous attribuons une fonction ou formulons une explication fonctionnelle, le succès ou l'échec de cette attribution/explication n'est ni arbitraire ni subjectif.

En généralisant à la téléologie l'argument auquel Dennett recourt pour l'intentionnalité, on dira que les fins et les fonctions qu'on attribue aux objets biologiques sont réelles et objectives (ce qui apparemment fait de nous des réalistes), bien qu'elles ne puissent être discernées que depuis le point de vue de celui qui adopte une certaine stratégie prédictive et que leur existence ne puisse être confirmée que par le succès de cette stratégie (ce qui apparemment fait de nous des interprétationnistes).

L'une des conséquences de cette posture est une indétermination radicale des fins et des fonctions. Cela veut dire que dans la plupart des cas nous sommes d'accord pour reconnaître qu'un trait (le cœur) a une fonction déterminée (pomper le sang), mais qu'en cas de conflit entre plusieurs interprétations concernant un cas particulier (le cœur d'un double accidentel, les membres des manchots), il n'existe pas nécessairement une unique réponse vraie à la question de sa fonctionnalité. Par conséquent, il n'existe pas non plus une unique conception valide des fonctions, ni une unique façon de les déterminer.

CONCLUSIONES GENERALES (ESPAÑOL)

« The main finding is that functions are relative to ends. »

Andrew Woodfield (1976, p. 110)

El concepto de función es teleológico y la teleología en biología es problemática. Sin embargo, los biólogos siguen recurriendo a ella. Uno de los retos del debate sobre las funciones, al que he dedicado la mayor parte de este trabajo, consiste en mostrar cómo las atribuciones y las explicaciones funcionales en biología pueden a pesar de todo ser legítimas y aceptables.

Una de las estrategias, propuesta por Robert Cummins (1975), trata de mostrar que el concepto no tiene implicaciones teleológicas. Según esta teoría, conocida como del « papel causal », las funciones no son más que disposiciones causales de los elementos de un sistema que contribuyen a explicar sus capacidades globales. Otras estrategias tratan por el contrario de justificar la teleología biológica al mostrar cómo se puede naturalizar. Una de ellas, inspirada por la cibernética, ofrece una explicación científica del comportamiento de los sistemas dirigidos hacia un fin. El carácter teleológico de las funciones se justifica en la medida en que los fines pueden ser descritos en términos mecánicos. Otra estrategia, propuesta inicialmente por Larry Wright (1973), interpreta las funciones en términos de selección. En el ámbito de la biología, esto significa que la función de un rasgo es aquello por lo que ha sido seleccionado (o tiene visos de serlo), o aquello por lo que contribuye (ahora, antes o después) a la *fitness* de los organismos que lo llevan. Desde esta perspectiva, atribuir una función a un rasgo es una forma de explicar su presencia mediante el mecanismo de la selección natural (o de otros mecanismos compatibles con la teoría darwiniana). Otros autores proponen aunar planteamientos diversos en una concepción unificada.

En cualquier caso, estas estrategias tratan de reducir o de traducir o de dar cuenta de las atribuciones y de las explicaciones funcionales en términos de atribuciones y explicaciones causales. En la práctica, esto significa que la mayoría de los autores proponen una definición del concepto de función formulada en términos de relaciones causales.

Las estrategias que he analizado son insatisfactorias e inútiles porque ofrecen soluciones para un problema inexistente. En efecto, las críticas que pretenden excluir la teleología de las ciencias naturales carecen de fundamento, pues no tienen necesariamente implicaciones ontológicas o metafísicas inaceptables. La finalidad no es una forma inválida de la causalidad ni una proyección de la psicología en los objetos naturales. No implica un compromiso con teorías y creencias contrarias a lo que nos enseñan las ciencias actuales, como el vitalismo, el animismo, el finalismo, el creacionismo, etc. De hecho, las explicaciones teleológicas se encuentran también en la física mediante principios variacionales (y no tienen ninguna implicación de ese tipo ya que los mismos fenómenos son explicables en términos no teleológicos).

Asimismo, he mostrado que son aceptables desde el puento de vista epistemológico. Por un lado, tienen fuerza explicativa y hacen posibles categorizaciones, inferencias y predicciones que, de hecho, contribuyen al avance de la biología. Por otro, son compatibles con teorías modernas de la explicación científica como la de James Woodward (2003), la cual es generalizable para abarcar tanto la finalidad como la causalidad.

La teleología es una forma de pensamiento aplicable a cualquier tipo de objeto que permite identificar lo que Woodward llama « relaciones invariantes de dependencia contrafáctica ». La finalidad es un tipo de relación contrafáctica invariante que vincula una acción (los medios) con su resultado (el fin) y una serie de constricciones (la situación, el contexto). De este modo, cuando se producen variaciones en los parámetros de la situación, los medios tienen que variar correlativamente para alcanzar el mismo fin. Por ejemplo, en óptica, esto significa que un rayo de luz modifica su trayectoria (los medios) siguiendo los índices de refracción de los medios que atraviesa (la situación) para minimizar el tiempo de recorrido (el fin). En biología, esto significa por ejemplo que un depredador elije su presa (los medios) según su valor calórico y el tiempo o el esfuerzo necesarios para su captura y su consumo (la situación) para maximizar el balance entre la energía conseguida y gastada (el fin).

Partiendo de un caso sencillo (el movimiento de un objeto sobre el plano) y variando las condiciones de su entorno (mediante obstáculos), pude comparar las explicaciones causales y teleológicas de un mismo comportamiento. Mostré que son diferentes y complementarias, porque identifican invariantes distintos, producen predicciones diferentes y no tienen el mismo grado de generalidad. Además, aunque siempre se pueden usar tanto unas como otras, su uso no es igualmente pertinetente en todas las situaciones, especialmente cuando son más complejas.

De manera general, el razonamiento teleológico se fundamenta en el principio de acción racional o eficiente. De acuerdo con éste, se espera de un agente que lleve a cabo la acción más eficiente — entre las que están disponibles— para alcanzar sus fines, teniendo en cuenta las condiciones de la situación. En otras palabras, en una acción dirigida hacia un fin, la relación entre los medios, los fines y las constricciones maximiza la eficiencia (la cual puede tomar diferentes formas: tiempo, dinero, esfuerzo, energía, etc.). Cuando una acción no se adecúa a ese principio, la interpretación teleológica no es aplicable. Al contrario, cuando lo hace, se puede inferir uno de los elementos de la relación mediante el conocimiento de los otros dos. El mismo principio se puede formular en términos de problemas–soluciones–condiciones.

No se trata de un principio ontológico ni metafísico, sino de una herramienta cognitiva que nos permite interpretar el mundo circundante para detectar regularidades interesantes y explotarlas con fines prácticos (escapar de un depredador, arreglar el coche, ganar al ajedrez) o teóricos (explicar un comportamiento animal, predecir un phenómeno físico). Según varios investigadores en psicología cognitiva, el razonamiento teleológico es un sistema de representación innato e independiente del razonamiento físico o mecánico, y tal vez anterior al razonamiento intencional. No se trata pues, como decía Jean Piaget, de un modo de razonamiento causal inmaduro. Tampoco es una tipo de mentalismo o de antropomorfismo, porque la atribución de un fin no implica la atribución de una mente ni de una forma u otra de intencionalidad.

Ciertamente, muchas explicaciones teleológicas son falsas, carentes de fundamento, e incluso absurdas o inaceptables desde el punto de vista científico. Es lo que ocurre en particular con las explicaciones religiosas de los fenómenos naturales. Por ejemplo, algunos creen que el Sol luce para alumbrarnos y que los terremotos se producen para castigar pecados. Ahora bien, es bastante fácil descartar este tipo de explicaciones con el principio de acción eficiente, un poco de espíritu crítico y algunos datos empíricos. El problema no está en el razonamiento teleológico como tal, sino en aquellos que lo usan, porque cometen los mismos errores de bulto con el razonamiento causal (pensamiento mágico, superstición, paraciencias, efecto cigüeña).

Si las explicaciones teleológicas en biología no son inválidas ni inaceptables, entonces el problema que las estrategias de naturalización tratan de resolver no existe. Por tanto, es posible definir el concepto de función no solamente en términos de relaciones causales (causas-efectos), sion también instrumentales (medios-fines). Como decía Woodfield en el epígrafe de estas conclusiones, el descubrimiento principal es que las funciones son relativas a fines. Atribuir una función a un rasgo significa que la actividad correspondiente es un medio de un fin del sistema al que pertenece.

Tratando de formular una definición teleológica del concepto de función tan general como fuera posible, llegué a la proposición siguiente :

> La/una función de un *token* de tipo X en un sistema de tipo S y en el contexto C es $f_{(x)}$ si:
>
> 1. $f_{(x)}$ es una consecuencia de X;
> 2. $f_{(x)}$ contribuye a un fin del sistema $F_{(S)}$;
> 3. $f_{(x)}$ pertenece a S.

Creo que esta definición teleológica permite evitar la mayor parte de las objeciones y contraejemplos habitualmente formulados contra las demás definiciones. Además, es suficientemente abstracta como para ser compatible (en la medida de lo posible) con muchas de esas definiciones entendidas como casos particulares de un caso más general. Por otra parte, pretende ser aplicable tanto a las funciones biológicas como a las de los artefactos.

Una atribución funcional es simultáneamente una explicación (no causal) de la presencia de un rasgo en un sistema determinado, a la manera de Wright, y una explicación (causal) de las capacidades o de los fines de ese mismo sistema, a la manera de Cummins.

Un aspecto importante de nuestra concepción es la distinción entre la arquitectura formal de un sistema (organismo o artefacto) y su arquitectura material. La primera se corresponde con las actividades u operaciones que son colectivamente constitutivas del sistema (esto es: metabolismo, circulación, respiración, etc.). La segunda se corresponde con la implementación material de esas operaciones, es decir con las estructuras concretas que las llevan a cabo. Ahora bien, una misma operación puede ser llevada a cabo por diferentes estructuras. Por tanto, no son las estructuras mismas las que tienen una función, sino lo que hacen.

Al atribuir una función a una operación, podemos explicar las *razones* de su presencia en el sistema, independientemente de las *causas* de la presencia de la estructura material que la lleva a cabo. Por ejemplo, podemos explicar racionalmente por qué el sistema circulatorio se compone de una bomba, independientemente del hecho que ésta sea un corazón natural o artificial, o un circuito de ECMO. Esto significa también que, desde mi punto de vista, y a diferencia de la concepción darwiniana, los órganos de los dobles accidentales son plenamente funcionales.

Por otra parte, el análisis funcional se entiende aquí como una ingeniería inversa que permite analizar un sistema, como el mecanismo de Anticitera, del cual desconocíamos de antemano tanto las capacidades como la utilidad y la historia causal, para intentar averiguar a la vez cómo podía funcionar y para qué podía servir teniendo en cuenta ciertos condicionantes como por ejemplo el contexto cultural de su construcción. Esto implica en particular que, a diferencia de la concepción de Cummins, los fines de un sistema no necesariamente son elegidos por el observador. Es la *organización* de un sistema, así como su contexto (en palabras de

Kitcher: el entorno, las presiones selectivas), quienes permiten la determinación de sus fines y de las funciones de sus partes. El criterio de optimalidad cumple un papel importante en esa determinación.

Otra manera de plantear el problema se conoce como el enfoque organizacional de las funciones (EOF), desarrollado de distintas maneras por varios autores. La propuesta que me parece más interesante es la de Mossio, Saborido, Moreno y sus colegas. Se apoya en una teoría científica de la organización de los seres vivos desde la cual pretenden naturalizar la teleología y la normatividad en biología. Muestran, como antes hiciera Kant, que la actividad de los seres vivos está orientada hacia un fin que no es extrínseco, sino autodeterminado por el sistema mismo. Este fin es el automantenimiento ou la autonomía. En ese marco, las funciones son una clase especial de relaciones causales propias de la organización de un sistema que, en la medida en que contribuyen al automantenimiento de ese sistema como conjunto, asimismo determinan etiológicamente las condiciones de existencia de los rasgos que las implementan.

Esta propuesta recuerda la de Rosenblueth, Wiener & Bigelow en su artículo precursor de la cibernética. Tanto unos como otros tratan de hallar una formulación científica rigurosa y novedosa para dar cuenta de la noción de finalidad. Si hubiera que traducir los fines y las funciones biológicas en términos de relaciones causales, el EOF sería sin duda la mejor de las alternativas. Sin embargo, creo que la teleología no necesita ser traducida de ese modo para verse justificada. Por otra parte, el EOF sólo da cuenta de una de las formas posibles de la teleología, la que atañe a la organización de lo vivo.

¿Tienen los rasgos biológicos una función determinada independientemente del observador? ¿Son las funciones una propiedad natural que podemos descubrir en el mundo (y acerca de la cual podemos estar equivocados) o una categoría humana que es relativa a nuestros intereses epistémicos y a nuestras estrategias interpretativas? Dos concepciones de las funciones pueden no coincidir sobre el hecho de que un rasgo posea o no una función, y a veces tampoco queda claro en el marco de una misma concepción. Desde una postura realista, como la que comparten los partidarios del enfoque etiológico y algunos partidarios del sistémico, la respuesta a la pregunta de si un rasgo tiene o no una función está en la naturaleza.

Las fonciones y la vida están íntimamente relacionadas, si tenemos en cuenta que los seres vivos (o los objetos de la biología, en sentido ámplio) son precisamente el tipo de cosas que se supone que tienen funciones, mientras que los demás sistemas naturales no. Por tanto, podríamos hallar respuestas relativas a las funciones en las investigaciones científicas sobre la vida misma, es decir sobre sus orígenes y sus mecanismos y condiciones mínimas de existencia. Esas investigaciones podrían por ejemplo ayudarnos a entender cómo aparecen las funciones cuando se produce la transición entre lo inerte y lo vivo, y en qué consisten esas

propiedades naturales que atribuimos a algunos objetos y a otros no. Para ello, es necesario poder distinguir entre unos y otros. Entre las preguntas que los investigadores se plantean está por un lado la de los criterios que permitirían identificar una forma de vida nueva y desconocida, bien sea en el fondo de una probeta o en un planeta lejano, y por otro lado la pregunta relativa a la definición universal de la vida y de lo vivo. Lo llamativo es que las estrategias con las que los investigadores tratan de contestar a estas preguntas son semejantes a las del debate sobre las funciones (sistémicas, darwinianas, mixtas), así como las dificultades con las que se encuentran, especialmente con el problema de establecer una distinción no ambigua entre la vida y la no-vida a partir de los criterios de complejidad o de selección natural. Además, las definiciones que proponen son a menudo funcionales, de modo que su aplicación depende de la interpretación del concepto de función.

Tal vez no haya una distinción entre la vida y la no-vida. Dicho de otro modo, tal vez dicha distinción no se encuentre en la naturaleza sino en el observador, es decir en nuestras categorías, nuestros modos de representación y nuestros métodos de investigación. Tal vez no haya una transición de lo inerte a lo vivo en la naturaleza misma, sino en la manera de describir los sistemas naturales a medida que se hacen más y más complejos y adquieren ciertas propiedades para el estudio de las cuales las herramientas de la biología son relevantes.

Si la vida y lo vivo fueran categorías humanas, eso no implicaría que son arbitrarias ni pondría en entredicho la cientificidad de la biología. Los planetas también son una categoría humana que tuvo que esperar hasta 2006 para ser definida por los astrónomos de manera estipulativa y convencional, y sólo para los planetas del Sistema Solar, pues los planetas extrasolares siguen carentes de definición. De este modo, los astrónomos no pretenden decir lo que *son* los planetas, sino que acuerdan emplear ese término para designar una clase de cuerpos cuyas propiedades les resultan interesantes. Los planetas son pues relativas a los intereses epistémicos de los astrónomos. A pesar de todo, ni los planetas son arbitrarios ni la astronomía deja de ser una ciencia rigurosa.

Si la vida no una propiedad natural, podría ser lo que Dennett llama una propiedad "encantadora". Aunque éstas sean relativas al observador, implican una forma de conocimiento del mundo externo, porque son la expresión de una relación entre nuestras estructuras cognitivas y las del mundo percibido, las cuales suelen ser la manifestación de estructuras más profundas. Esta relación hace posibles las generalizaciones inductivas, las predicciones exitosas y las explicaciones científicas. Hay ahí un conocimiento objetivo que no requiere un compromiso ontológico respecto de la cosa conocida.

Lo anterior me lleva a pensar que quizás las funciones tampoco sean propiedades naturales en el sentido en que las entienden Millikan, Longy y Davies. Al contrario, defiendo que las funciones son un tipo de relación

propia de la actitud teleológica, del mismo modo que la causalidad es un tipo de relación propia de la actitud física. Ambas actitudes son modos innatos, autónomos y complementarios de representación de la realidad que pueden *de jure* y *de facto* ser empleadas en el ámbito científico. Aunque no existan independientemente de nuestros intereses epistémicos y de nuestras estrategias predictivas, las causas y las funciones pueden a pesar de todo ser objetivas, porque el éxito de esas estrategias no depende de nosotros. En otras palabras, cuando atribuimos una función o formulamos una explicación funcional, el éxito o el fracaso de esa atribución/explicación no es arbitrario ni subjetivo.

Si empleamos para la teleología el argumento que Dennett emplea para la intencionalidad, podemos decir que los fines y las funciones que atribuimos a los objetos biológicos son reales y objetivos (lo cual nos convierte, al menos aparentemente, en realistas), aunque sólo puedan ser discernidos desde el punto de vista de aquel que adopta una determinada estrategia predictiva y que su existencia sólo pueda ser confirmada mediante el éxito de esa estrategia (lo cual nos convierte, al menos aparentemente, en interpretacionistas).

Una de las consecuencias de esta postura es una indeterminación radical de los fines y de las funciones. Significa que en la mayoría de los casos estamos de acuerdo en que cierto rasgo (el corazón) tiene una función determinada (bombear sangre), pero cuando hay desacuerdo entre varias interpretaciones de un caso particular (el corazón de los dobles accidentales, los miembros de los pingüinos), no existe necesariamente una única respuesta verdadera para la pregunta de su función. Por tanto, tampoco existe una única concepción válida de las funciones, ni una única forma de identificarlas.

RÉFÉRENCES

Abkarian, M., Massiera, G., Berry, L., Roques, M., & Braun-Breton, C. (2011). A novel mechanism for egress of malarial parasites from red blood cells. *Blood*, *117*(15), 4118-4124. https://doi.org/10.1182/blood-2010-08-299883

Abrams, M. (2005). Teleosemantics Without Natural Selection. *Biology & Philosophy*, *20*(1), 97-116. https://doi.org/10.1007/s10539-005-0359-7

Abrams, M. (2007). Fitness and propensity's annulment? *Biology and Philosophy*, *22*(1).

Abrams, P. (2001). Adaptationism, optimality models and tests of adaptive scenarios. In S. H. Orzack & E. Sober (Éd.), *Adaptationism and Optimality* (p. 273–302). Cambridge University Press.

Achinstein, P. (1977). Function Statements. *Philosophy of Science*, *44*(3), 341-367.

Achinstein, P. (1978). Teleology and Mentalism. *The Journal of Philosophy*, *75*(10), 551-553.

Achinstein, P. (1983). *The Nature of Explanation*. Oxford; New York: Oxford University Press.

Adami, C. (2002). What is complexity? *BioEssays*, *24*(12), 1085–1094.

Adami, C., Ofria, C., & Collier, T. C. (2000). Evolution of biological complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(9), 4463.

Adams, F. R. (1979). A Goal-State Theory of Function Attributions. *Canadian Journal of Philosophy*, *9*(3), 493-518.

Agassi, J. (2008). *Science and Its History: A Reassessment of the Historiography of Science*. Springer Science & Business Media.

Agazzi, E. (1988). L'objectivité scientifique. In E. Agazzi (Éd.), *L'objectivité dans les différentes sciences* (p. 13-25). Fribourg: Editions Universitaires de Fribourg.

Allen, C. (2002). Real Traits, Real Functions? In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 373-389). Oxford; New York: Oxford University Press.

Allen, C. (2003). Teleological Notions in Biology. In F. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University.

Allen, C., & Bekoff, M. (1995a). Biological Function, Adaptation, and Natural Design. *Philosophy of Science*, *62*(4), 609-622.

Allen, C., & Bekoff, M. (1995b). Function, Natural Design, and Animal Behavior: Philosophical and Ethological Considerations. In N. S. Thompson (Éd.), *Perspectives in Ethology: Volume 11: Behavioral Design* (p. 1-47). New York: Plenum Press.

Allen, C., & Bekoff, M. (1995c). Teleology, function, design, and the evolution of animal behaviour. *Trends in Ecology and Evolution*, *10*(6), 253-255.

Amundson, R. (2000). Against normal function. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 33-53. https://doi.org/10.1016/S1369-8486(99)00033-3

Amundson, R., & Lauder, G. V. (1994). Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology. *Biology and Philosophy*, *9*(4), 443-469.

Anglada-Escudé, G., Amado, P. J., Barnes, J., Berdiñas, Z. M., Butler, R. P., Coleman, G. A. L., … Zechmeister, M. (2016). A terrestrial planet candidate in a temperate orbit around Proxima Centauri. *Nature*, *536*(7617), 437-440. https://doi.org/10.1038/nature19106

Ariew, A. (2002). Platonic and Aristotelian Roots of Teleological Arguments in Cosmology and Biology. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 7-32). Oxford; New York: Oxford University Press.

Ariew, A., Cummins, R., & Perlman, M. (Éd.). (2002). *Functions : new essays in the philosophy of psychology and biology*. Oxford; New York: Oxford University Press.

Ariew, A., & Lewontin, R. C. (2004). The confusions of fitness. *British Journal for the Philosophy of Science*, *55*(2), 347-363.

Arp, R., & Smith, B. (2008). Function, Role, and Disposition in Basic Formal Ontology. *Nature Precedings*, (713). https://doi.org/10.1038/npre.2008.1941.1

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., & Claverie, J.-M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences*, *108*(42), 17486-17491. https://doi.org/10.1073/pnas.1110889108

Ashby, W. R. (1956). *An introduction to cybernetics*. New York, J. Wiley. Consulté à l'adresse http://archive.org/details/introductiontocy00ashb

Asher, Y. M., & Kemler Nelson, D. G. (2008). Was it designed to do that? Children's focus on intended function in their conceptualization of artifacts. *Cognition*, *106*(1), 474–483.

Atlan, H. (1979). *Entre le cristal et la fumée. Essai sur l'organisation du vivant*. Paris: Seuil.

Atran, S. (1995). Causal constraints on categories and categorical constraints on biological reasoning across cultures. In D. Sperber, D. Premack, & A. J. Premack (Éd.), *Causal cognition: A multidisciplinary debate*. New York: Clarendon Press; Oxford University Press.

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and brain sciences*, *27*(06), 713–730.

Avise, J. C., Bowen, B. W., Lamb, T., Meylan, A. B., & Bermingham, E. (1992). Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Mol. Biol. Evol*, *9*(3), 457–473.

Ayala, F. J. (1970). Teleological Explanations in Evolutionary Biology. *Philosophy of Science*, *37*(1), 1-15. https://doi.org/10.2307/186024

Ayala, F. J. (1972). The Autonomy of Biology as a Natural Science. In A. D. Breck & W. Yourgrau (Éd.), *Biology, History, and Natural Philosophy* (p. 1-16). Boston, MA: Springer US. Consulté à l'adresse http://link.springer.com/10.1007/978-1-4684-1695-4_1

Ayala, F. J. (1977). Teleological Explanations. In T. Dobzhansky (Éd.), *Evolution* (p. 497-504). San Francisco: W.H. Freeman & Company.

Bachelard, G. (1938). *La formation de l'esprit scientifique*. Paris: Vrin.

Bachmann, P. A., Walde, P., Luisi, P. L., & Lang, J. (1990). Self-replicating reverse micelles and chemical autopoiesis. *Journal of the American Chemical Society*, *112*(22), 8200–8201.

Baillargeon, R., & DeVos, J. (1991). Object Permanence in Young Infants: Further Evidence. *Child Development*, *62*(6), 1227-1246.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191-208.
Ball, P. (2011). A metaphor too far. *Nature News*. https://doi.org/10.1038/news.2011.115
Barahona, A., & Torrens, E. (2004). El « telos » aristotélico y su influencia en la biología moderna. *Ludus Vitalis: Revista de Filosofia de las Ciencias de la Vida*, *12*(21), 161-178.
Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8
Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in cognitive sciences*, *4*(1), 29–34.
Basri, G., & Brown, M. E. (2006). Planetesimals to brown dwarfs: What is a planet? *Annual Review of Earth and Planetary Sciences*, *34*, 193-216.
Battro, A. M. (1966). *Dictionnaire d'épistémologie génétique*. Paris: Presses universitaires de France.
Beatty, J. (1980). Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology. *Philosophy of Science*, *47*(4), 532-561. https://doi.org/10.1086/288955
Beckner, M. (1959). *The biological way of thought*. NY: COLUMBIA UNIV PR. Consulté à l'adresse http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=phl&AN=PHL1015647&site=ehost-live&scope=site
Beckner, M. (1969). Function and Teleology. *Journal of the History of Biology*, *2*(1), 151-164. https://doi.org/10.1007/BF00137271
Bedau, M. A. (1990). Against Mentalism in Teleology. *American Philosophical Quarterly*, *27*(1), 61-70.
Bedau, M. A. (1991). Can Biological Teleology be Naturalized? *The Journal of Philosophy*, *88*, 647-655.
Bedau, M. A. (1992a). Goal-directed systems and the good. *Monist*, *75*(1), 34-52.
Bedau, M. A. (1992b). Where's the Good in Teleology? *Philosophy and Phenomenological Research*, *52*(4), 781-806.
Bedau, M. A. (1996). The Nature of Life. In M. Boden (Éd.), *The Philosophy of Artificial Life* (p. 332-357). Oxford University Press.
Bedau, M. A. (1998). Four Puzzles about Life. *Artificial Intelligence*, *4*, 125-140.
Bedau, M. A. (2007). Une vue fonctionnelle de la cellule vivante minimale. In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 5-13). Vuibert.
Bedau, M. A., & Packard, N. H. (1992). Measurement of Evolutionary Activity, Teleology, and Life. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Éd.), *Artificial Life II, Santa Fe Institute Studies in the Sciences of Complexity* (Vol. 10, p. 431-461). Redwood City, CA: Addison-Wesley.
Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology*, *41*(2), 328.
Benner, S. A., & Sismour, M. A. (2005). Synthetic Biology. *Nature*, *6*, 533-543.
Bensaude-Vincent, B., & Newman, W. R. (Éd.). (2007). *The artificial and the natural: an evolving polarity*. Cambridge, Mass: MIT Press.
Berger, L. R., Hawks, J., de Ruiter, D. J., Churchill, S. E., Schmid, P., Delezene, L. K., … others. (2015). Homo naledi, a new species of the genus Homo from the Dinaledi Chamber, South Africa. *eLife*, *4*, e09560.
Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale*. Paris: Flammarion.
Bersini, H., & Reisse, J. (Éd.). (2007). *Comment définir la vie?* Vuibert.

Bertalanffy, L. von. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, *1*, 134-165.
Bertalanffy, L. von. (1968). *General system theory: Foundations development applications*. Harmondsworth, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord: Penguin books.
Bigelow, J., & Pargetter, R. (1987). Functions. *The Journal of Philosophy*, *84*, 181-196.
Birch, J. (2015). Natural selection and the maximization of fitness. *Biological Reviews*. https://doi.org/10.1111/brv.12190
Bird, A., & Tobin, E. (2017). Natural Kinds. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Metaphysics Research Lab, Stanford University. Consulté à l'adresse https://plato.stanford.edu/archives/spr2017/entries/natural-kinds/
Biro, S. (2013). The role of the efficiency of novel actions in infants' goal anticipation. *Journal of experimental child psychology*, *116*(2), 415–427.
Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: development through cue-based bootstrapping. *Developmental science*, *10*(3), 379–398.
Bitbol, M. (1996). *Mécanique quantique: Une introduction philosophique*. Paris: Flammarion.
Bitbol, M. (1998). *L'Aveuglante proximité du réel*. Paris: Flammarion.
Bitbol, M. (2000). *Physique et philosophie de l'esprit*. Paris: Flammarion.
Bitbol, M., & Luisi, P. L. (2004). Autopoiesis with or without cognition: defining life at its edge. *Journal of The Royal Society Interface*, *1*(1), 99-107. https://doi.org/10.1098/rsif.2004.0012
Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567.
Blay, M. (2003). *Grand Dictionnaire de la Philosophie*. Larousse; CNRS Editions.
Blay, M., & Halleux, R. (Éd.). (1998). *La science classique: XVIe-XVIIIe siècle. Dictionnaire critique*. Paris: Flammarion.
Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, *60*, 1-29.
Bloom, P. (2004). *Descartes' baby : how the science of child development explains what makes us human*. New York: Basic Books.
Bombelli, P., Howe, C. J., & Bertocchini, F. (2017). Polyethylene bio-degradation by caterpillars of the wax moth Galleria mellonella. *Current Biology*, *27*(8), R292-R293. https://doi.org/10.1016/j.cub.2017.02.060
Boorse, C. (1976). Wright on Functions. *Philosophical Review*, *85*, 70-86.
Boorse, C. (1977). Health as a Theoretical Concept. *Philosophy of Science*, *44*(4), 542-573.
Boorse, C. (2002). A Rebuttal on Functions. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 63-112). Oxford; New York: Oxford University Press.
Borges, J. L. (1944). Ficciones. In *Obras Completas* (Vol. 1, p. 455-567). Buenos Aires: Emecé.
Borges, J. L. (1952). Otras inquisiciones. In *Obras completas* (Vol. 2, p. 368-453). Buenos Aires: Emecé.
Borgonie, G., Garcia-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., … Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. *Nature*, *474*(7349), 79-82. https://doi.org/10.1038/nature09974
Born, M. (1939). Cause, Purpose and Economy of Natural Laws Minimum Principles in Physics. *Nature*, *143*(3618), 357-361. https://doi.org/10.1038/143357a0
Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. OUP Oxford.
Bouchard, F. (2013). How ecosystem evolution strengthens the case for function pluralism. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 83-95). Springer.

Boyd, R. (1999). Kinds, Complexity, and Multiple Realization. *Philosophical Studies*, *95*(1-2), 67–98.
Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., … Raoult, D. (2011). Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proceedings of the National Academy of Sciences*, *108*(25), 10296-10301. https://doi.org/10.1073/pnas.1101118108
Brack, A. (2007). Comment définir la vie? In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 15-23). Paris: Vuibert.
Braillard, P.-A., & Malaterre, C. (2015). *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Springer.
Braithwaite, R. B. (1946). Teleological Explanation: The Presidential Address. *Proceedings of the Aristotelian Society*, *47*, i-xx. https://doi.org/10.2307/4544417
Braithwaite, R. B. (1953). *Scientific Explanation*. Cambridge University Press.
Brandon, R. N. (1978). Adaptation and Evolutionary Theory. *Studies In History and Philosophy of Science*, *9*, 181-206.
Brandon, R. N. (1981). Biological Teleology: Questions and Explanations. *Studies in the History and Philosophy of Science*, *12*(2), 91-105.
Brandon, R. N. (2013). A general case for functional pluralism. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 97-104). Springer.
Braude, S. (1997). The predictive power of evolutionary biology and the discovery of eusociality in the naked mole rat. *Reports of the National Center for Science Education*, *17*(4), 12–15.
Britt, R. R. B. | A. (2006, août 16). Nine Planets Become 12 with Controversial New Definition. Consulté 5 avril 2017, à l'adresse http://www.space.com/2741-planets-12-controversial-definition.html
Brown, M. E. (2006). A band of astronomers at the IAU meeting in Prague are revolting against the proposed IAU definition. Consulté 3 avril 2017, à l'adresse http://web.gps.caltech.edu/~mbrown/whatsaplanet/revolt.html
Brown, M. E. (2008, juin 22). What's in a name? Consulté 10 avril 2017, à l'adresse http://www.mikebrownsplanets.com/2008/06/whats-in-name.html
Brown, M. E. (2010). *How I killed Pluto and why it had it coming* (1st ed). New York: Spiegel & Grau.
Brown, T. L. (2003). *Making truth: metaphor in science*. Urbana: University of Illinois Press.
Bruylands, G., Bartik, K., & Reisse, J. (2010). Is it useful to have a clear-cut definition of life? On the use of fuzzy-logic in prebiotic chemistry. *Origins of Life and Evolution of Biospheres*, (40), 137-143.
Buller, D. J. (1998). Etiological Theories of Function: A Geographical Survey. *Biology and Philosophy*, *13*, 505-527.
Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking & Reasoning*, *15*(2), 167–196.
Busino, G. (1988). L'objectivité dans les sciences humaines. In E. Agazzi (Éd.), *L'objectivité dans les différentes sciences* (p. 179-186). Fribourg: Editions Universitaires de Fribourg.
Cairns-Smith, A. G., & Hartman, H. (1986). *Clay minerals and the origin of life*. Cambridge University Press.
Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences*, *12*(5), 187–192.

Camara, C. G., Escobar, J. V., Hird, J. R., & Putterman, S. J. (2008). Correlation between nanosecond X-ray flashes and stick–slip friction in peeling tape. *Nature*, *455*(7216), 1089-1092. https://doi.org/10.1038/nature07378

Cameron, R. J. (2000). *Teleology in Aristotle and Contemporary Philosophy of Biology: An account of the Nature of Life* (Thèse de Doctorat). University of Colorado.

Campbell, J. (2008). Interventionism, control variables and causation in the qualitative world. *Philosophical Issues*, *18*(1), 426–445.

Canfield, J. (1964). Teleological explanation in biology. *The British Journal for the Philosophy of Science*, *14*(56), 285-95.

Canfield, J. (1965). Teleological Explanation in Biology: A Reply. *The British Journal for the Philosophy of Science*, *15*(60), 327-331.

Canfield, J. (Éd.). (1966). *Purpose in nature*. New Jersey: Prentice-Hall.

Canguilhem, G. (1965). *La Connaissance de la vie*. Paris: Vrin.

Canguilhem, G. (1966). *Le normal et le pathologique*. Paris: PUF.

Cannon, W. B. (1929). Organisation for physiological homeostasis. *Physiological Reviews*, *9*(3), 399-431.

Caponi, G. (2001a). Biología funcional vs. biología evolutiva. *Episteme, Porto Alegre*, *12*, 23-46.

Caponi, G. (2001b). Función y adaptación: dos modos de la teleología. *Epistemologia e Historia de la Ciencia*, *7*(7), 66-70.

Carey, S. (1985). *Conceptual change in childhood*. MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language*, (3), 167-181.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carrara, M., & Vermaas, P. E. (2009). The Fine-Grained Metaphysics of Artifactual and Biological Functional Kinds. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, *169*(1), 125-143.

Cartwright, N. (1986). Two kinds of teleological explanation. In B. Donagan, A. Perovich, & M. Wedin (Éd.), *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her Seventy-Fifth Birthday* (p. 201-210). Dordrecht: Reidel.

Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science*, *8*(6), 472–480.

Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition*, *103*(1), 120–130.

Casler, K., & Kelemen, D. (2008). Developmental continuity in teleo-functional explanation: Reasoning about nature among Romanian Romani adults. *Journal of Cognition and Development*, *9*(3), 340–362.

Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., … Wyrzykowski, Ł. (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature*, *481*(7380), 167-169. https://doi.org/10.1038/nature10684

Cavicchioli, R. (2002). Extremophiles and the Search for Extraterrestrial Life. *Astrobiology*, *2*(3), 281-292. https://doi.org/10.1089/153110702762027862

Chao, L. (2000). The Meaning of Life. *BioScience*, *50*(3), 245-250.

Check, E. (2005). Synthetic biology: Designs on life. *Nature*, *438*(7067), 417-418. https://doi.org/10.1038/438417a

Christensen, W. (2012). Natural Sources of Normativity. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 104-112.

Christensen, W. D. (1996). A complex systems theory of teleology. *Biology and Philosophy*, *11*(3), 301-320.
Christensen, W. D., & Bickhard, M. H. (2002). The Process Dynamics of Normative Function. *Monist: An International Quarterly Journal of General Philosophical Inquiry*, *85*(1), 3-28.
Chyba, C. F., & McDonald, G. D. (1995). The Origin of Life in the Solar System: Current Issues. *Annu. Rev. Earth Planet. Sci.*, *23*, 215-249.
Clair, J. J. H. S., & Rutz, C. (2013). New Caledonian crows attend to multiple functional properties of complex tools. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1630), 20120415. https://doi.org/10.1098/rstb.2012.0415
Claverie, J.-M. (2005). Un virus encore plus géant que les autres. *M/S : médecine sciences*, *21*(1), 15-16.
Cleland, C. E., & Chyba, C. F. (2002). Defining « Life ». *Origins of Life and Evolution of Biospheres*, *32*(4), 387-393.
Cohen, A. A. (2007). A Kantian Stance on Teleology in Biology. *South African Journal of Philosophy*, *26*(2), 109-121.
Cohen, J. (1951). Teleological Explanation. *Proceedings of the Aristotelian Society*, *51*, 255-292. https://doi.org/10.2307/4544486
Collier, J. (2000). Autonomy and Process Closure as the Basis for Functionality. *Annals of the New York Academy of Sciences*, *901*(1), 280-290. https://doi.org/10.1111/j.1749-6632.2000.tb06287.x
Collins, A. W. (1978). Teleological Reasoning. *The Journal of Philosophy*, *75*(10), 540-550.
Couchman, J. R., & Rees, D. A. (1979). The behaviour of fibroblasts migrating from chick heart explants: changes in adhesion, locomotion and growth, and in the distribution of actomyosin and fibronectin. *Journal of Cell Science*, *39*(1), 149-165.
Creamer, J. S., Mora, M. F., & Willis, P. A. (2017). Enhanced Resolution of Chiral Amino Acids with Capillary Electrophoresis for Biosignature Detection in Extraterrestrial Samples. *Analytical Chemistry*, *89*(2), 1329-1337. https://doi.org/10.1021/acs.analchem.6b04338
Csibra, G. (2003). Teleological and Referential Understanding of Action in Infancy. *Philosophical Transactions: Biological Sciences*, *358*(1431), 447-458. https://doi.org/10.2307/3558125
Csibra, G. (2005). Mirror neurons and action observation. Is simulation involved. *What do mirror neurons mean*. Consulté à l'adresse http://www.cbcd.bbk.ac.uk/people/scientificstaff/gergo/pub/index.html/pub/mirror.pdf
Csibra, G. (2008a). Action mirroring and action understanding: an alternative account. *Sensorymotor Foundations of Higher Cognition. Attention and Performance XXII*, 435–459.
Csibra, G. (2008b). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*(2), 705–717.
Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111.
Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255.
Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI*, 249–274.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153.
Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason'in infancy. *Cognition*, *72*(3), 237–267.
Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, *72*, 741-765.
Cummins, R. (2002a). Haugeland on Representation and Intentionality. In H. Clapin (Éd.), *Philosophy of Mental Representation* (p. 122-137). Oxford: Clarendon Press.
Cummins, R. (2002b). Neo-Teleology. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 157-172). Oxford; New York: Oxford University Press.
Daeschler, E. B., Shubin, N. H., & Jenkins, F. A. (2006). A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature*, *440*(7085), 757-763. https://doi.org/10.1038/nature04639
Damiano, L., & Luisi, P. L. (2010). Towards an autopoietic redefinition of life. *Origins of Life and Evolution of Biospheres*, (40), 145-9.
Danto, A. C. (1965). *Narration and Knowledge (including the integral text of Analytical Philosophy of History)*. New York: Columbia University Press.
Darwin, C. (1872). *The origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life.* (6e éd.). London: John Murray.
Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, *60*(23), 685–700.
Davies, J. (2010). Anthropomorphism in science. *EMBO Reports*, *11*(10), 721. https://doi.org/10.1038/embor.2010.143
Davies, P. S. (2000a). Malfunctions. *Biology and Philosophy*, *15*(1), 19-38.
Davies, P. S. (2000b). The Nature of Natural Norms: Why Selected Functions Are Systemic Capacity Functions. *Noûs*, *34*(1), 85-107.
Davies, P. S. (2001). *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, MA: Bradford Books.
Davies, P. S. (2009). Conceptual Conservatism: The Case of Normative Functions. In U. Krohs & P. Kroes (Éd.), *Functions in Biological and Artificial Worlds* (p. 127-145). Cambridge, MA; London: MIT Press.
Dawkins, R. (1976). *The Selfish Gene* (30th anniversary ed). Oxford: Oxford University Press.
Dawkins, R. (1986). *The Blind Watchmaker*. New York: Norton.
Dawkins, R. (1998). Accumulating small chance. In M. Ruse (Éd.), *Philosophy of Biology* (p. 62-68). New York: Prometheus Books.
Dawkins, R. (2016). *The extended phenotype: The long reach of the gene*. Oxford University Press. Consulté à l'adresse https://books.google.es/books?hl=en&lr=&id=kOvmDAAAQBAJ&oi=fnd&pg=PP1&dq=extended+phenotype+example&ots=1Fmz6DUaeY&sig=nyTj3cI9TRcErrMyIkJnrl-pc-o
de Lorenzo, V., & Danchin, A. (2008). Synthetic biology: discovering new worlds and new words. The new and not so new aspects of this emerging research field. *EMBO reports*, *9*(9), 822-827. https://doi.org/10.1038/embor.2008.159
Defeyter, M. A., & German, T. P. (2003). Acquiring an understanding of design: evidence from children's insight problem solving. *Cognition*, *89*(2), 133-155. https://doi.org/10.1016/S0010-0277(03)00098-2
Dennett, D. C. (1971). Intentional Systems. *The Journal of Philosophy*, *68*(4), 87-106. https://doi.org/10.2307/2025382
Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1990). The Interpretation of Texts, People and Other Artifacts. *Philosophy and Phenomenological Research*, *50*, 177-194.
Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, *88*(1), 27-51. https://doi.org/10.2307/2027085
Dennett, D. C. (1993). *La Conscience expliquée*. Paris: Odile Jacob.
Dennett, D. C. (1995). *Darwin's Dangerous Idea*. London: Penguin.
Dennett, D. C. (1996). *Kinds of minds: toward an understanding of consciousness*. New York: BasicBooks.
Dennett, D. C. (2000). Postmodernism and truth. In *The Proceedings of the Twentieth World Congress of Philosophy* (Vol. 8, p. 93–103).
Denton, M. (1998). Behond the reach of chance. In M. Ruse (Éd.), *Philosophy of Biology* (p. 47-61). New York: Prometheus Books.
Deplazes, A., & Huppenbauer, M. (2009). Synthetic organisms and living machines. *Systems and Synthetic Biology*, *3*(1-4), 55-63. https://doi.org/10.1007/s11693-009-9029-4
Devitt, M. (2011). Natural kinds and biological realisms. *Carving Nature at Its Joints: Natural Kinds in Metaphysics and Science*, 155–170.
DiGregorio, B. E. (2000, juillet 16). Viking Data May Hide New Evidence For Life. *Space Daily*. Consulté à l'adresse http://www.spacedaily.com/news/mars-life-00g.html
Doolittle, W. F. (2013). Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences*, *110*(14), 5294-5300. https://doi.org/10.1073/pnas.1221376110
Doolittle, W. F., Brunet, T. D. P., Linquist, S., & Gregory, T. R. (2014). Distinguishing between "Function" and "Effect" in Genome Biology. *Genome Biology and Evolution*, *6*(5), 1234-1237. https://doi.org/10.1093/gbe/evu098
Dorais, L.-J. (2011). Les mots inuits pour la neige et la glace. In *L'Encyclopédie Canadienne*. Toronto: Historica Canada. Consulté à l'adresse http://www.thecanadianencyclopedia.ca/fr/article/inuit-words-for-snow-and-ice/
Drake, N. (2011). Subterranean worms from hell. *Nature*. https://doi.org/10.1038/news.2011.342
Dretske, F. (1986). Misrepresentation. Consulté à l'adresse http://philpapers.org/rec/DREM
Dretske, F. (1988). *Explaining behavior: reasons in a world of causes* (5th print). Cambridge, Mass.: MIT Press.
Ducasse, C. J. (1925). Explanation, Mechanism, and Teleology. *The Journal of Philosophy*, *22*, 150-154.
Ducasse, C. J. (1959). Life, Telism, and Mechanism. *Philosophy and Phenomenological Research*, *20*(1), 18-24. https://doi.org/10.2307/2104950
Dyson, F. (1985). *Origins of Life*. Cambridge University Press.
Edmonds, B. (1999). What is complexity? In F. Heylighen, J. Bollen, & A. Riegler (Éd.), *The evolution of complexity* (p. 1-16). Springer.
Edwards, A. W. F. (2007). Maximisation principles in evolutionary biology. In M. Matthen & C. Stephens (Éd.), *Philosophy of Biology* (p. 335–349). Amsterdam: Elsevier.
Egri, Á., Blahó, M., Kriska, G., Farkas, R., Gyurkovszky, M., Åkesson, S., & Horváth, G. (2012). Polarotactic tabanids find striped patterns with brightness and/or polarization modulation least attractive: an advantage of zebra stripes. *The Journal of Experimental Biology*, *215*(5), 736-745. https://doi.org/10.1242/jeb.065540
Ehrman, L., & Grosseield, J. (1980). What is Natural, What is Not? *Hastings Center Report*, *10*(5), 10-11. https://doi.org/10.2307/3561042

Elbaz, D. (2007). Combien d'étoiles contient notre galaxie ? Combien de galaxies dans l'Univers exploré ? *La Recherche*, (412), 95.
Elder, C. L. (2007). On the place of artifacts in ontology. In E. Margolis & S. Laurence (Éd.), *Creations of the Mind: Theories of Artifacts and Their Representation* (p. 33-51). Oxford University Press.
Elliott, T. A., Linquist, S., & Gregory, T. R. (2014). Conceptual and Empirical Challenges of Ascribing Functions to Transposable Elements. *The American Naturalist*, *184*(1), 14-24. https://doi.org/10.1086/676588
Enç, B. (1979). Fonction Attributions and Functional Explanations. *Philosophy of Science*, *46*, 343-365.
Enç, B. (2002). Indeterminacy of Function Attributions. In A. Ariew (Éd.), *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford Univ Pr.
Enç, B., & Adams, F. R. (1992). Functions and Goal Directedness. *Philosophy of Science*, *59*(4), 635-654.
Endy, D. (2005). Foundations for engineering biology. *Nature*, *438*(7067), 449-453. https://doi.org/10.1038/nature04342
ESA. (2017). Searching for signs of life on Mars. Consulté 3 avril 2017, à l'adresse http://exploration.esa.int/mars/43608-life-on-mars/
Eshuis, R., Coventry, K. R., & Vulchanova, M. (2009). Predictive eye movements are driven by goals, not by the mirror neuron system. *Psychological Science*, *20*(4), 438–440.
Espagnat, B. d'. (1994). *Le réel voilé: Analyse des concepts quantiques*. Paris: Fayard.
Etxeberria, A. (2006). Organismo y organización en la biología teórica: ¿Vuelta al organicismo? *Ludus Vitalis*, *14*(26), 3-38.
Fagot-Largeault, A. (1995). Le vivant. In D. Kambouchner (Éd.), *Notions de philosophie* (Vol. 1, p. 231-300). Paris: Gallimard.
Fagot-Largeault, A. (2002). L'ordre vivant. In D. Andler, A. Fagot-Largeault, & B. Saint-Sernin (Éd.), *Philosophie des sciences* (Vol. 1, p. 483-575). Paris: Gallimard.
Ferguson, K. G. (2007). Biological Function and Normativity. *Philo: A Journal of Philosophy*, *10*(1), 17-26.
Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*(6), 939-946. https://doi.org/10.1177/0956797612464058
Ferrater Mora, J. (1986). Función. *Diccionario de filosofía* (Vol. 2, p. 1300). Madrid: Alianza.
Fisher, R. A. (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.
Fitzpatrick, W. J. (2000). *Teleology and the norms of nature*. New York: Garland Publishing.
FitzSimmons, N. N., Moritz, C., & Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution*, *12*(3), 432.
Flatow, I. (2006, septembre 8). Astronomers Prepare to Fight Pluto Demotion. *Talk of the Nation: Science Friday*. npr. Consulté à l'adresse http://www.npr.org/templates/story/story.php?storyId=5788798
Fleischaker, G. R. (1990). Origins of life: An operational definition. *Origins of Life and Evolution of Biospheres*, *20*(2), 127-137.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R. A., Henrissat, B., … others. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, *336*(6089), 1715–1719.

Flourens, P. (1857). *Histoire de la découverte de la circulation du sang* (2e éd.). Paris: Garnier frères. Consulté à l'adresse http://gallica.bnf.fr/ark:/12148/bpt6k28154h

Forster, A. C., & Church, G. M. (2006). Towards synthesis of a minimal cell. *Molecular Systems Biology*, *2*. https://doi.org/10.1038/msb4100090

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research*, *117*(1), 5-16. https://doi.org/10.1016/j.virusres.2006.01.010

Forterre, P. (2010). Defining Life: The Virus Viewpoint. *Origins of Life and Evolution of Biospheres*, *40*, 151-160. https://doi.org/10.1007/s11084-010-9194-1

Foucault, M. (1966). *Les Mots et les choses : Une archéologie des sciences humaines*. Paris: Gallimard.

Frankfurt, H. G., & Poole, B. (1966). Functional Analyses in Biology. *The British Journal for the Philosophy of Science*, *17*(1), 69-72.

Freeth, T. (2008). Calendars with Olympiad display and eclipse prediction on the Antikythera Mechanism. *Nature*, *454*, 614-617. https://doi.org/doi:10.1038/nature07130

Freeth, T., Bitsakis, Y., Moussas, X., Seiradakis, J. H., Tselikas, A., Mangou, H., … Edmunds, M. G. (2006). Decoding the ancient Greek astronomical calculator known as the Antikythera Mechanism. *Nature*, *444*(7119), 587-591. https://doi.org/10.1038/nature05357

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative Function Demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, *117*(1), 1-8. https://doi.org/10.1016/j.cognition.2010.06.003

Galison, P. (1987). *How Experiments End*. Chicago: University of Chicago Press.

Gánti, T. (2003). *The Principles of Life*. Oxford University Press.

Gardner, A. (2009). Adaptation as organism design. *Biology Letters*, *5*(6), 861-864. https://doi.org/10.1098/rsbl.2009.0674

Garson, J. (2016). *A Critical Overview of Biological Functions*. Springer.

Garvey, B. (2007). *Philosophy of biology*. Stocksfield: Acumen. Consulté à l'adresse http://dx.doi.org/10.1017/UPO9781844653812

Gayon, J. (1993). La biologie: entre loi et histoire. *Philosophie*, *38*.

Gayon, J. (2010a). Defining Life: Synthesis and Conclusions. *Origins of Life and Evolution of Biospheres*, *40*(2), 231-244. https://doi.org/10.1007/s11084-010-9204-3

Gayon, J. (2010b). Raisonnement fonctionnel et niveaux d'organisation en biologie. In J. Gayon & A. de Ricqlès (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 125-138). Paris: PUF.

Gayon, J., & de Ricqlès, A. (Éd.). (2010). *Les fonctions: des organismes aux artefacts*. Paris: PUF.

Gergely, G. (2011). Kinds of agents. In U. C. Goswami (Éd.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., p. 76-105). Chichester ; Malden, MA: Wiley-Blackwell.

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. *Cognition*, *63*(2), 227-233. https://doi.org/10.1016/S0010-0277(97)00004-8

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287-292.

Gergely, G., Nádasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193.

Germain, P.-L., Ratti, E., & Boem, F. (2014). Junk or Functional DNA? ENCODE and the Function Controversy. *Biology and Philosophy*, *29*(6), 807–831.
German, T. P., & Johnson, S. C. (2002). Functions and the Origins of the Design Stance. *Journal of Cognition and Development*, *3*(3), 279-300.
Gibbs, W. W. (2004). Synthetic life. *Scientific American*, *Mai 2004*, 74-81.
Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., … Smith, H. O. (2008). Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science*, *319*(5867), 1215-1220. https://doi.org/10.1126/science.1151721
Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., … Venter, J. C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, *329*(5987), 52-56. https://doi.org/10.1126/science.1190719
Gillon, M., Triaud, A. H. M. J., Demory, B.-O., Jehin, E., Agol, E., Deck, K. M., … Queloz, D. (2017). Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature*, *542*(7642), 456-460. https://doi.org/10.1038/nature21360
Ginsborg, H. (2006). Kant's Biological Teleology and Its Philosophical Significance. In *A Companion to Kant*. Malden MA: Blackwell Publishing. Consulté à l'adresse http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=phl&AN=PHL2086839&site=ehost-live&scope=site
Ginsborg, H. (2014). Oughts Without Intentions: A Kantian Approach to Biological Functions. In E. Watkins & I. Goy (Éd.), *Kant's Theory of Biology* (p. 259–274). De Gruyter.
Girotto, V., Pievani, T., & Vallortigara, G. (2014). Supernatural beliefs: Adaptations for social life or byproducts of cognitive adaptations? *Behaviour*, *151*, 385–402.
Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., … Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 425-430. https://doi.org/10.1073/pnas.0510013103
Godfrey-Smith, P. (1993). Functions: Consensus without unity. In *The Philosophy of Biology* (p. 258-279). Oxford University Press.
Godfrey-Smith, P. (1994). A Modern History Theory of Functions. *Noûs*, *28*(3), 344-362.
Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. H. Orzack & E. Sober (Éd.), *Adaptationism and optimality* (p. 335–357). Cambridge University Press.
Godfrey-Smith, P. (2004). Mental Representation, Naturalism, and Teleosemantics. In D. Papineau & G. MacDonald (Éd.), *Teleosemantics: New Philosophical Essays*. Oxford University Press.
Godfrey-Smith, P. (2007). Information in biology. In D. L. Hull & M. E. Ruse (Éd.), *The Cambridge Companion to the Philosophy of Biology* (p. 103-119). Cambridge: Cambridge University Press.
Gogarten, J. P. (2011, juillet). *Molecular evolution before LUCA and the rooted Net of Life*. Comunication présenté à Origins 2011: ISSOL - The International Astronomy Society and Bioastronomy (IAU C51) Joint Conference, Montpellier. Consulté à l'adresse http://www.exobiologie.fr/origins2011/slides/
Goldman, A. (2011, juillet). *Viruses aren't Life, but neither are you*. Comunication présenté à Origins 2011: ISSOL - The International Astronomy Society and Bioastronomy (IAU C51) Joint Conference, Montpellier. Consulté à l'adresse http://www.exobiologie.fr/origins2011/slides/
Goldman, A. I. (2012). Theory of mind. *The Oxford handbook of philosophy of cognitive science*, 402–424.

Golomb, B. A. (2013). Lab life: Chocolate habits of Nobel prizewinners. *Nature*, *499*(7459), 409-409. https://doi.org/10.1038/499409a

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants. *Pediatrics*, *56*(4), 544-549.

Gould, J. L. (2007). Animal Artifacts. In E. Margolis & S. Laurence (Éd.), *Creations of the Mind: Theories of Artifacts and Their Representation* (p. 249–266). Oxford University Press.

Gould, S. J. (1993). La rueda de la fortuna y la cuña del progreso. In L. Preta (Éd.), *Imágenes y metáforas de la ciencia* (p. 59-73). Madrid: Alianza.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *205*(1161, The Evolution), 581-598.

Gould, S. J., & Vrba, E. (1982). Exaptations: A Missing Term in the Science of Form. *Paleobiology*, *8*, 5-15.

Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., Zufall, R. A., & Elhaik, E. (2013). On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE. *Genome Biology and Evolution*, *5*(3), 578-590. https://doi.org/10.1093/gbe/evt028

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science*, *17*(6), 455–459.

Grene, M., & Depew, D. (2004). *The philosophy of biology : an episodic history*. New York: Cambridge University Press.

Griffiths, P. E. (1993). Functional Analysis and Proper Functions. *The British Journal for the Philosophy of Science*, *44*(3), 409-422.

Griffiths, P. E. (1994). Cladistic classification and functional explanation. *Philosophy of Science*, *61*(2), 206-227.

Griffiths, P. E. (1996). The historical turn in the study of adaptation. *The British journal for the philosophy of science*, *47*(4), 511.

Griffiths, P. E. (2001). Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science*, *68*(3), 394-412.

Griffiths, P. E. (2006). Function, Homology, and Character Individuation. *Philosophy of science*, *73*(1), 1–25.

Gruner, R. (1966). Teleological and functional explanations. *Mind: A Quarterly Review of Philosophy*, *75*, 516-526.

Gutheil, G., Bloom, P., Valderrama, N., & Freedman, R. (2004). The role of historical intuitions in children's and adults' naming of artifacts. *Cognition*, *91*(1), 23-42.

Gutheil, G., Vera, A., & Keil, F. C. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition*, *66*, 33-49.

Hackett, J. (2015, février 13). Pluto's ongoing identity crisis stirs planet definition debate » Scienceline. Consulté 8 avril 2017, à l'adresse http://scienceline.org/2015/02/plutos-ongoing-identity-crisis-stirs-planet-definition-debate/

Haddock, S. H., Moline, M. A., & Case, J. F. (2010). Bioluminescence in the sea. *Marine Science*, *2*. Consulté à l'adresse http://www.annualreviews.org/eprint/pbWcm4DyqvQGJ7F2qGkc/full/10.1146/annurev-marine-120308-081028

Häggqvist, S. (2013). Teleosemantics: Etiological Foundations. *Philosophy Compass*, *8*(1), 73-83. https://doi.org/10.1111/j.1747-9991.2012.00532.x
Haldane, J. (1949). What is life? In M. Ruse (Éd.), *Philosophy of Biology* (p. 32-34). Amherst, NY: Prometheus Books.
Halleux, R. (1998). Harvey. In M. Blay & R. Halleux (Éd.), *La Science classique XVIe-XVIIe siècle. Dictionnaire critique* (p. 270-276). Paris: Flammarion.
Hanke, D. (2004). Teleology: the explanation that bedevils biology. In J. Cornwell (Éd.), *Explanations: Styles of Explanation in Science*. Oxford University Press.
Harari, Y. N. (2015). *Sapiens: a brief history of humankind* (First U.S. edition). New York: Harper.
Hardcastle, V. G. (2002). On the Normativity of Functions. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 144-156). Oxford; New York: Oxford University Press.
Hare, B., & Woods, V. (2013). *The genius of dogs. How dogs are smarter than you think*. New York, New York: Dutton, a member of Penguin Group (USA) Inc.
Harris, E. E. (1959). Teleology and Teleological Explanation. *The Journal of Philosophy*, *56*(1), 5-25. https://doi.org/10.2307/2022800
Harrison, T. (2006). *Harrison : principios de medicina interna* (16e éd.). México: McGraw-Hill.
Hauser, M. D. (2002). *À quoi pensent les animaux?* Paris: Odile Jacob.
Hauser, M. D., & Wood, J. (2010). Evolving the capacity to understand actions, intentions, and goals. *Annual review of psychology*, *61*, 303–324.
Hegde, N. R., Maddur, M. S., Kaveri, S. V., & Bayry, J. (2009). Reasons to include viruses in the tree of life. *Nat Rev Micro*, *7*(8), 615. https://doi.org/10.1038/nrmicro2108-c1
Hempel, C. G. (1965). The Logic of Funcional Analysis. In *Aspects of Scientific Explanation*. New York: Free Press.
Henshilwood, C. S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A. T., … Wintle, A. G. (2002). Emergence of Modern Human Behavior: Middle Stone Age Engravings from South Africa. *Science*, *295*(5558), 1278-1280. https://doi.org/10.1126/science.1067575
Hernik, M., & Csibra, G. (2009). Functional understanding facilitates learning about tools in human children. *Current Opinion in Neurobiology*, *19*(1), 34-38. https://doi.org/10.1016/j.conb.2009.05.003
Herrel, A., Huyghe, K., Vanhooydonck, B., Backeljau, T., Breugelmans, K., Grbac, I., … Irschick, D. J. (2008). Rapid large-scale evolutionary divergence in morphology and performance associated with exploitation of a different dietary resource. *Proceedings of the National Academy of Sciences*, *105*(12), 4792-4795. https://doi.org/10.1073/pnas.0711998105
Herrmann, E., Wobber, V., & Call, J. (2008). Great apes'(Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus) understanding of tool functional properties after limited experience. *Journal of Comparative Psychology*, *122*(2), 220.
Hill, K., Kaplan, H., Hawkes, K., & Hurtado, A. M. (1987). Foraging decisions among Ache hunter-gatherers: new data and implications for optimal foraging models. *Ethology and Sociobiology*, *8*(1), 1–36.
Hilpinen, R. (1993). Authors and Artifacts. *Proceedings of the Aristotelian Society*, *93*, 155-178.
Hilpinen, R. (2011). Artifact. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Winter 2011). Metaphysics Research Lab, Stanford University. Consulté à l'adresse https://plato.stanford.edu/archives/win2011/entries/artifact/
Holmes, B. (2005, février 12). Alive! The race to create life from scratch. *New Scientist*, (2486).

Holmes, E. C. (2011). What Does Virus Evolution Tell Us about Virus Origins? *J. Virol.*, *85*(11), 5247-5251. https://doi.org/10.1128/JVI.02203-10
Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal cognition*, *8*(3), 164–181.
Horseshoe crab. (2010). Consulté 5 avril 2010, à l'adresse http://www.mbl.edu/marine_org/marine_org.php?func=detail&myID=BX151
Huneman, P. (Éd.). (2013a). *Functions: selection and mechanisms*. Dordrecht: Springer Netherlands. Consulté à l'adresse http://link.springer.com/10.1007/978-94-007-5304-4
Huneman, P. (2013b). Introduction. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 1-16). Springer.
Huneman, P. (2013c). Weak Realism in the Etiological Theory of Functions. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 105–130). Springer.
Hyde, D. (2008). Sorites Paradox. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Consulté à l'adresse http://plato.stanford.edu/entries/sorites-paradox/
IAU. (2006, août 24). General assembly: results of the IAU resolution votes. Consulté 3 avril 2017, à l'adresse https://www.iau.org/news/pressreleases/detail/iau0603/
IAU definition of planet. (2010). In *Wikipedia*. Consulté à l'adresse http://en.wikipedia.org/wiki/IAU_definition_of_planet
Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York: Psychology Press.
Inagaki, K., & Hatano, G. (2006). Young Children's Conception of the Biological World. *Current Directions in Psychological Science*, *15*(4), 177-181.
Introduction to the Myxini. (2010). Consulté 5 avril 2010, à l'adresse http://www.ucmp.berkeley.edu/vertebrates/basalfish/myxini.html
Jackson, F. (1996). The coevolutionary relationship of humans and domesticated plants. *American Journal of Physical Anthropology*, *101*(S23), 161-176. https://doi.org/10.1002/(SICI)1096-8644(1996)23+<161::AID-AJPA6>3.0.CO;2-8
Jacob, F. (1970). *La logique du vivant: une histoire de l'hérédité*. Paris: Gallimard.
Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166.
Jacob, F. (2000). Qu'est-ce que la vie? In Y. Michaud (Éd.), *Université de tous les savoirs: Qu'est-ce que la vie?* (Vol. 1, p. 23-36). Paris: Odile Jacob.
Jaffe, S. (2005). In the Business of Synthetic Life. *Scientific American*, *292*(4), 40-41.
Jax, K. (2005). Function and "functioning" in ecology: what does it mean? *Oikos*, *111*(3), 641–648.
Johnson, S. C., & Ok, S.-J. (2007). Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, *22*(3), 310–322.
Johnson, S. S., Hebsgaard, M. B., Christensen, T. R., Mastepanov, M., Nielsen, R., Munch, K., … Willerslev, E. (2007). Ancient bacteria show evidence of DNA repair. *Proceedings of the National Academy of Sciences*, *104*(36), 14401-14405. https://doi.org/10.1073/pnas.0706787104
Joordens, J. C. A., d'Errico, F., Wesselingh, F. P., Munro, S., de Vos, J., Wallinga, J., … Roebroeks, W. (2014). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature*, *518*(7538), 228-231. https://doi.org/10.1038/nature13962
Joyce, G. F. (1989). RNA evolution and the origins of life. *Nature*, *338*(6212), 217–224.
Joyce, G. F. (1994). Foreword. In *Origins of Life: The Central Concepts* (p. xi-xii). Boston: Jones & Bartlett.

Joyce, G. F. (2002). The antiquity of RNA-based evolution. *Nature*, *418*(6894), 214–221.
Kant, I. (1790). *Critique de la faculté de juger*. (A. Philonenko, Trad.) (1993-Ed. rev. avec des notes nouv éd.). Paris: Vrin.
Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Éd.), *Modularity and constraints in language and cognition* (Vol. 25, p. 103-138). Erlbaum.
Keil, F. C. (1995). The growth of causal understanding of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Éd.), *Causal cognition: A multidisciplinary debate* (p. 234-262). Clarendon Press.
Keil, F. C. (2006). Explanation and Understanding. *Annual review of psychology*, *57*, 227-254. https://doi.org/10.1146/annurev.psych.57.102904.190100
Keil, F. C. (2013). The roots of folk biology. *Proceedings of the National Academy of Sciences*, *110*(40), 15857-15858. https://doi.org/10.1073/pnas.1315113110
Kelemen, D. (1999a). Function, goals and intention: children's teleological reasoning about objects. *Trends in Cognitive Sciences*, *3*(12), 461-468.
Kelemen, D. (1999b). The scope of teleological thinking in preschool children. *Cognition*, *70*, 241-272.
Kelemen, D. (1999c). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology*, *35*(6), 1440-1452.
Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. *Creations of the mind: Theories of artifacts and their representation*, 212–230.
Kelemen, D., & DiYanni, C. (2005). Intuitions about origins: Purpose and intelligent design in children's reasoning about nature. *Journal of Cognition and Development*, *6*(1), 3–31.
Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143.
Kelemen, D., Rottman, J., & Seston, R. (2012). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*. Consulté à l'adresse http://psycnet.apa.org/journals/xge/142/4/1074/
Kelemen, D., Widdowson, D., Posner, T., Brown, A. L., & Casler, K. (2003). Teleo-functional constraints on preschool children's reasoning about living things. *Developmental Science*, *6*(3), 329–345.
Keller, E. F. (2000). Decoding the genetic program. In P. J. Beurton, R. Falk, & H.-J. Rheinberger (Éd.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press.
Keller, E. F. (2003). *Making sense of life: explaining biological development with models, metaphors, and machines*. Cambridge, Mass.: Harvard Univ. Press.
Keller, E. F. (2007). Comment définir la vie? In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 45-50). Vuibert.
Kellis, M., Wold, B., Snyder, M. P., Bernstein, B. E., Kundaje, A., Marinov, G. K., … Hardison, R. C. (2014). Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences*, *111*(17), 6131-6138. https://doi.org/10.1073/pnas.1318948111
Kemler Nelson, D. G., Egan, L. C., & Holt, M. B. (2004). When children ask,"What is it?" what do they want to know about artifacts? *Psychological Science*, *15*(6), 384–389.

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition*, *77*(2), 133–168.
Kemler Nelson, D. G., Herron, L., & Holt, M. B. (2003). The sources of young children's name innovations for novel artifacts. *Journal of child language*, *30*(04), 823–843.
Kemler Nelson, D. G., Herron, L., & Morris, C. (2002). How children and adults name broken objects: Inferences and reasoning about design intentions in the categorization of artifacts. *Journal of Cognition and Development*, *3*(3), 301–332.
Kemler Nelson, D. G., O'Neil, K. A., & Asher, Y. M. (2008). A mutually facilitative relationship between learning names and learning concepts in preschool children: The case of artifacts. *Journal of Cognition and Development*, *9*(2), 171–193.
Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-Year-Olds Will Name Artifacts by Their Functions. *Child Development*, *71*(5), 1271-1288.
Kermen, F., Midroit, M., Kuczewski, N., Forest, J., Thévenet, M., Sacquet, J., … Mandairon, N. (2016). Topographical representation of odor hedonics in the olfactory bulb. *Nature Neuroscience*, *19*(7), 876-878. https://doi.org/10.1038/nn.4317
Kinoshita, S., Kageyama, S., Iba, K., Yamada, Y., & Okada, H. (1975). Utilization of a cyclic dimer and linear oligomers of ε-aminocaproic acid by Achromobacter guttatus KI72. *Agricultural and Biological Chemistry*, *39*(6), 1219-1223. https://doi.org/10.1271/bbb1961.39.1219
Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and Cognition*, *12*(4), 752–769.
Kiritani, O. (2011). Function and Modality. *Journal of Mind and Behavior*, *32*(1), 1-4.
Kistler, M. (2013). La réduction, l'émergence, l'unité de la science et les niveaux de réalité. In M. Silberstein (Éd.), *Matériaux philosophiques et scientifiques pour un matérialisme contemporain* (Vol. 1). Éditions matériologiques.
Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science*, *48*(4), 507-531.
Kitcher, P. (1984). 1953 and all that. A Tale of Two Sciences. *The Philosophical Review*, *93*(3), 335-373.
Kitcher, P. (1985). Two Approaches to Explanation. *The Journal of Philosophy*, *82*(11), 632-639.
Kitcher, P. (1993). Function and Design. In D. L. Hull & M. Ruse (Éd.), *The Philosophy of Biology* (p. 258-279). Oxford University Press.
Klarsfeld, A., & Revah, F. (2000). *Biologie de la mort*. Paris: Odile Jacob.
Klump, B. C., Wal, J. E. M. van der, Clair, J. J. H. S., & Rutz, C. (2015). Context-dependent 'safekeeping' of foraging tools in New Caledonian crows. *Proc. R. Soc. B*, *282*(1808), 20150278. https://doi.org/10.1098/rspb.2015.0278
Kool, E. T., Morales, J. C., & Guckian, K. M. (2000). Mimicking the structure and function of DNA: insights into DNA stability and replication. *Angewandte Chemie International Edition*, *39*(6), 990–1009.
Kramer, P. J. (1998). Misuse of the term strategy. In M. Ruse (Éd.), *Philosophy of Biology* (p. 185-186). New York: Prometheus Books.
Krimbas, C. B. (2004). On fitness. *Biology and Philosophy*, *19*(2), 185-203.
Krohs, U. (2008). Co-Designing Social Systems by Designing Technical Artifacts: A Conceptual Approach. In P. E. Vermaas, P. Kroes, A. Light, & S. A. Moore (Éd.), *Philosophy and Design: From Engineering to Architecture*. Dordrecht: Springer. Consulté à l'adresse http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=phl&AN=PHL2123785&site=ehost-live&scope=site

Krohs, U. (2010). Dys-, mal- et non- : L'autre face de la fonctionnalité. In J. Gayon & A. de Ricqlès (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 337-52). Paris: PUF.
Krohs, U., & Kroes, P. (Éd.). (2009). *Functions in Biological and Artificial Worlds*. Cambridge Mass.: MIT Press.
Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition*, *94*(1), 95-103.
Kuhn, T. S. (1957). *The Copernican Revolution*. Harvard University Press.
Kuhn, T. S. (1983). *La structure des révolutions scientifiques*. Paris: Flammarion.
Kupiec, J.-J. (2009). L'ADN entre hasard et contraintes. *Pour la Science*, *385*, 88-95.
Küppers, B.-O. (1995). The context-dependance of biological information. *Ludus Vitalis*, *3*(5), 5-17.
La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., … Raoult, D. (2003). A Giant Virus in Amoebae. *Science*, *299*(5615), 2033. https://doi.org/10.1126/science.1081867
La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., … Raoult, D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature*, *455*(7209), 100-104. https://doi.org/10.1038/nature07218
Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 451.
Lakoff, G. (1987). *Women, Fire and Dangerous Things, What Categories Reveal About the Mind*. Chicago; London: University of Chicago Press.
Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: University of Chicago Press.
Lange, M. (2004). The Autonomy of Functional Biology: A Reply to Rosenberg. *Biology and Philosophy*, *19*(1), 93-109.
Larison, B., Harrigan, R. J., Thomassen, H. A., Rubenstein, D. I., Chan-Golston, A. M., Li, E., & Smith, T. B. (2015). How the zebra got its stripes: a problem with too many solutions. *Royal Society Open Science*, *2*(1), 140452. https://doi.org/10.1098/rsos.140452
Lawler, D., & Encabo, J. V. (2011). Clases artificiales. *Azafea: Revista de Filosofía*, *12*(1), 119-147.
Lazcano, A. (2007). Vers une définition évolutionniste de la vie: implications concernant l'origine des systèmes vivants. In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 51-56).
Lazcano, A. (2008). What is Life? A Brief Historical Overview. *Chemistry and Biodiversity*, *5*, 1-15.
Lederberg, J. (1960). Exobiology: Approaches to Life beyond the Earth. *Science*, *132*(3424), 393-400.
Lee, D. H., Severin, K., & Ghadiri, M. R. (1997). Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. *Current Opinion in Chemical Biology*, *1*(4), 491-496. https://doi.org/10.1016/S1367-5931(97)80043-9
Legendre, M., Arslan, D., Abergel, C., & Claverie, J.-M. (2012). Genomics of Megavirus and the elusive fourth domain of Life. *Communicative & Integrative Biology*, *5*(1), 102-106.
Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., … Claverie, J.-M. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences*, 201320670. https://doi.org/10.1073/pnas.1320670111
Lehman, H. S. (1965a). Functional Explanation in Biology. *Philosophy of Science*, *32*(1), 1-20.
Lehman, H. S. (1965b). Teleological Explanation in Biology. *The British Journal for the Philosophy of Science*, *15*(60), 327.

Lennox, J. G. (2010). La fonction biologique: phylogénie d'un concept. In J. Gayon & A. de Ricqlès (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 17-42). Paris: PUF.
Lenski, R. E., Ofria, C., Collier, T. C., & Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, *400*(6745), 661–664.
Lenton, T. M. (1998). Gaia and natural selection. *Nature*, *394*(6692), 439-447. https://doi.org/10.1038/28792
Leslie, A. M. (1982). The perception of causality in infants. *Perception*, *11*(2), 173–186.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288.
Levitis, D. A., Lidicker, W. Z., & Freund, G. (2009). Behavioural biologists don't agree on what constitutes behaviour. *Animal behaviour*, *78*(1), 103-110. https://doi.org/10.1016/j.anbehav.2009.03.018
Lewens, T. (2000). Function talk and the artefact model. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 95-111. https://doi.org/10.1016/S1369-8486(99)00040-0
Lewens, T. (2002). Adaptationism and Engineering. *Biology and Philosophy*, *17*(1), 1-31.
Lewens, T. (2004). *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge MA: Bradford Book/MIT Pr.
Lewens, T. (2007). Adaptation. In D. L. Hull & M. Ruse (Éd.), *The Cambridge Companion to the Philosophy of Biology* (p. 1-21). New York: Cambridge University Press.
Lewontin, R. C. (2001a). In the Beginning Was the Word. *Science*, *291*(5507), 1263-1264. https://doi.org/10.1126/science.1057124
Lewontin, R. C. (2001b). *The Triple Helix: Gene, Organism, and Environment*. Harvard University Press.
Lichnerowicz, A. (1987). Mathématique et physique. *Bulletin de la Classe des sciences. Académie royale de Belgique*, *5*(73), 95-112.
Lincoln, T. A., & Joyce, G. F. (2009). Self-Sustained Replication of an RNA Enzyme. *Science*, *323*(5918), 1229-1232. https://doi.org/10.1126/science.1167856
Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, *160*, 35-42. https://doi.org/10.1016/j.cognition.2016.12.007
Lloyd, S. (2001). Measures of complexity: a nonexhaustive list. *IEEE Control Systems*, *21*(4), 7-8. https://doi.org/10.1109/MCS.2001.939938
Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464-470.
Lombrozo, T. (2009). Explanation and categorization: How "why?" informs "what?" *Cognition*, *110*(2), 248–253.
Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332.
Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, *99*, 167-204.
Longy, F. (2009). How Biological, Cultural, and Intended Functions Combine. In U. Krohs & P. Kroes (Éd.), *Functions in Biological and Artificial Worlds* (p. 51-67). Cambridge Mass.: MIT Press.
Longy, F. (2010). Ce qu'explique une explication fonctionnelle. In A. de Ricqlès & J. Gayon (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 405-418). PUF.

Longy, F. (2013). Artifacts and Organisms: A Case for a New Etiological Theory of Functions. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 185–211). Springer.
Longy, F. (2015). Biological Functions and Semantic Contents: The Teleosemantics. In T. Heams, P. Huneman, G. Lecointre, & M. Silberstein (Éd.), *Handbook of Evolutionary Thinking in the Sciences* (p. 853-877). Dordrecht: Springer Netherlands. Consulté à l'adresse http://link.springer.com/10.1007/978-94-017-9014-7_40
López García, P. (2007). Comment définir la vie? In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 57-64).
Loveland-Curtze, J., Miteva, V. I., & Brenchley, J. E. (2009). Herminiimonas glaciei sp. nov., a novel ultramicrobacterium from 3042 m deep Greenland glacial ice. *International Journal of Systematic and Evolutionary Microbiology*, *59*(6), 1272-1277. https://doi.org/10.1099/ijs.0.001685-0
Lovelock, J. (1995). *The ages of Gaia: a biography of our living earth*. Oxford University Press.
Lovelock, J., & Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis. *Month*, 2–10.
Luisi, P. L. (1998). About various definitions of life. *Origins of Life and Evolution of the Biosphere*, *28*, 613-622.
Luisi, P. L. (2002). Toward the engineering of minimal living cells. *The Anatomical Record*, *268*(3), 208-214. https://doi.org/10.1002/ar.10155
Luisi, P. L. (2003). Autopoiesis: a review and a reapraisal. *Naturwissenschaften*, *90*(2), 49-59.
Luisi, P. L. (2006). *The Emergence of Life: From Chemical Origins to Synthetic Biology*. Cambridge University Press.
Luisi, P. L. (2007). Chemical aspects of synthetic biology. *Chemistry and Biodiversity*, *4*(4), 603.
Luisi, P. L., Chiarabelli, C., & Stano, P. (2006). From Never Born Proteins to Minimal Living Cells: Two Projects in Synthetic Biology. *Origins of Life and Evolution of the Biosphere*, *36*, 605-616.
Luisi, P. L., Ferri, F., & Stano, P. (2006). Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften*, *93*, 1-13.
Luisi, P. L., Oberholzer, T., & Lazcano, A. (2002). The Notion of a DNA Minimal Cell: A General Discourse and Some Guidelines for an Experimental Approach. *Helvetica Chimica Acta*, *85*(6), 1759-1777. https://doi.org/10.1002/1522-2675(200206)85:6<1759::AID-HLCA1759>3.0.CO;2-7
Luisi, P. L., & Varela, F. J. (1989). Self-replicating micelles — A chemical version of a minimal autopoietic system. *Origins of Life and Evolution of Biospheres*, *19*(6), 633-643. https://doi.org/10.1007/BF01808123
Luisi, P. L., Walde, P., & Oberholzer, T. (1999). Lipid vesicles as possible intermediates in the origin of life. *Current Opinion in Colloid & Interface Science*, *4*(1), 33–39.
Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453-460.
Luo, Y., & Baillargeon, R. (2005). When the ordinary seems unexpected: evidence for incremental physical knowledge in young infants. *Cognition*, *95*(3), 297-328.
Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive psychology*, *58*(4), 441–486.
Lurz, R. (2009). Animal Minds. *Internet Encyclopedia of Philosophy*. Consulté à l'adresse http://www.iep.utm.edu/ani-mind/#SSH1ciii

Macdonald, G., & Papineau, D. (Éd.). (2006). *Teleosemantics: new philosophical essays*. Oxford ; New York: Clarendon Press.
Malaterre, C. (2010). On what it is to fly can tell us something about what it is to live. *Origins of Life and Evolution of Biospheres*, (40), 169-77.
Marchant, J. (2006). In search of lost time. *Nature*, *444*(7119), 534-538. https://doi.org/10.1038/444534a
Marcos, A. (2009). Funciones en biología: una perspectiva aristotélica. *Dialogo Filosofico*, *25:2*(74), 231-248.
Margoliash, D., & Tchernichovski, O. (2015). Marmoset kids actually listen. *Science*, *349*(6249), 688-689. https://doi.org/10.1126/science.aac7860
Margolis, E., & Laurence, S. (2014). Concepts. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). Consulté à l'adresse http://plato.stanford.edu/archives/spr2014/entries/concepts/
Margot, J.-L. (2015). A Quantitative Criterion for Defining Planets. *The Astronomical Journal*, *150*(6), 185. https://doi.org/10.1088/0004-6256/150/6/185
Margulis, L. (1981). Symbiosis in cell evolution: Life and its environment on the early earth, 419.
Margulis, L., & Sagan, D. (1995). *What is Life?* New York: Simon and Schuster.
Markushin, Y. (2010, janvier 10). Psychology in chess: is it really there? Consulté 23 août 2015, à l'adresse http://www.thechessworld.com/learn-chess/4-healthpsycology/112-psychology-in-chess-is-it-really-there
Matthen, M. (1988). Biological functions and perceptual content. *The Journal of Philosophy*, *85*(1), 5–27.
Matthews, R. (2000). Storks deliver babies (p= 0.008). *Teaching Statistics*, *22*(2), 36–38.
Maturana, H., & Varela, F. J. (1973). *De Máquinas y Seres Vivos* (6ème). Santiago: Editorial Universitaria.
Maund, B. (2000). Proper functions and Aristotelian functions in biology. *Studies In History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 155-178. https://doi.org/10.1016/S1369-8486(99)00038-2
Mawhin, J. (1998). Le principe de moindre action et la finalité. *Revue d'éthique et de théologie morale, « Le supplément »*, (205), 49-82.
Maynard Smith, J. (1986). *The problems of biology*. Oxford: Oxford University Press.
Maynard Smith, J. (2000). The Concept of Information in Biology. *Philosophy of Science*, *67*, 177-194.
Maynard Smith, J. (2006). Optimization theory in evolution. In E. Sober (Éd.), *Conceptual issues in evolutionary biology* (3e éd.). Cambridge, Mass: MIT Press.
Maynard Smith, J., & Szathmáry, E. (1999). *Les origines de la vie*. Paris: Dunod.
Mayr, E. (1961). Cause and Effect in Biology. *Science*, *134*(3489), 1501-1506.
Mayr, E. (1974). Teleological and teleonomic: A new analysis. *Boston studies in the philosophy of science*, *14*(1). Consulté à l'adresse http://evolution.freehostia.com/wp-content/uploads/2007/07/mayr_1974_teleological_and_teleonomic.rtf
Mayr, E. (1988). *Towards a New Philosophy of Biology*. Cambridge, MA: Harvard University Press.
Mayr, E. (1992). The Idea of Teleology. *Journal of the History of Ideas*, *53*(1), 117-135.
Mayr, E. (1996). The Autonomy of Biology: The Position of Biology Among the Sciences. *Quarterly Review of Biology*, *71*(1), 97-106.
Mayr, E. (1997). *Qu'est-ce que la biologie?* Paris: Fayard.

Mayr, E. (2007). *What Makes Biology Unique?: Considerations on the Autonomy of a Scientific Discipline*. Cambridge University Press.
McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18.
McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3-and 5-year-old children. *Cognitive Development*, *22*(3), 353–364.
McLaughlin, P. (2001). *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. New York: Cambridge University Press.
McLaughlin, P. (2009). Functions and Norms. In U. Krohs & P. Kroes (Éd.), *Functions in Biological and Artificial Worlds* (p. 93-102). Cambridge Mass.: MIT Press.
McNamara, J. M., Houston, A. I., & Collins, E. J. (2001). Optimality models in behavioral biology. *Siam Review*, *43*(3), 413–466.
McShea, D. W. (2000). Functional Complexity in Organisms: Parts as Proxies. *Biology and Philosophy*, *15*(5), 641-668. https://doi.org/10.1023/A:1006695908715
Méary, D., Kitromilides, E., Mazens, K., Graff, C., & Gentaz, E. (2007). Four-day-old human neonates look longer at non-biological motions of a single point-of-light. Consulté à l'adresse http://dx.plos.org/10.1371/journal.pone.0000186
Mesli, M. (2013). La fabuleuse histoire du principe de moindre action: de Fermat à Feynman. Université de Toulon: Canal-U. Consulté à l'adresse https://www.canal-u.tv/video/universite_de_toulon/la_fabuleuse_histoire_du_principe_de_moindre_action_thinsp_de_fermat_a_feynman.16464
Messerli, F. H. (2012). Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine*, *367*(16), 1562-1564. https://doi.org/10.1056/NEJMon1211064
Metzger, P. (2015, avril 13). Nine Reasons Why Pluto Is a Planet. Consulté 3 avril 2017, à l'adresse http://www.philipmetzger.com/blog/nine-reasons-why-pluto-is-a-planet/
Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science*, *117*(3046), 528–529.
Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
Millikan, R. G. (1989a). An ambiguity in the notion "function". *Biology and Philosophy*, *4*(2), 172-176. https://doi.org/10.1007/BF00127747
Millikan, R. G. (1989b). In Defense of Proper Functions. *Philosophy of Science*, *56*(2), 288-302.
Millikan, R. G. (1993). Propensities, Exaptations and the Brain. In R. G. Millikan (Éd.), *White Queen Psychology and Other Essays for Alice* (p. 31-50). Cambridge, MA: MIT Press.
Millikan, R. G. (1999). Historical Kinds and the « Special Sciences ». *Philosophical Studies*, *95*(1-2), 45–65.
Millikan, R. G. (2000). Reading Mother Nature's Mind. In D. Ross, A. Brook, & D. Thompson (Éd.), *Dennett's Philosophy: A comprehensive Assessment* (p. 55-76). Cambridge, Mass.: MIT Press.
Millikan, R. G. (2002). Biofunctions: Two paradigms. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 113-143). Oxford; New York: Oxford University Press.

Mills, C. M., & Keil, F. C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, *87*(1), 1-32. https://doi.org/10.1016/j.jecp.2003.09.003

Mills, S. K., & Beatty, J. H. (1979). The Propensity Interpretation of Fitness. *Philosophy of Science*, *46*(2), 263-286.

Mitchell, M. (2009). *Complexity: A Guided Tour*. New York: Oxford University Press.

Mitchell, S. D. (1993). Dispositions or Etiologies? A Comment on Bigelow and Pargetter. *Journal of Philosophy*, *90*, 249-259.

Mitchell, S. D. (1995). Function, fitness and disposition. *Biology and Philosophy*, *10*(1), 39-54. https://doi.org/10.1007/BF00851986

Mix, L. J. (2015). Defending definitions of life. *Astrobiology*, *15*(1), 15-19.

Molina Pérez, A. (2006). Objetividad versus inteligibilidad de las funciones biológicas: La paradoja normativa y el autismo epistemológico de las ciencias modernas. *Ludus Vitalis: Revista de Filosofia de las Ciencias de la Vida*, *14*(26), 39-67.

Molina Pérez, A. (2009). Techniques et concepts du vivant en biologie synthétique. *Ludus Vitalis: Revista de Filosofia de las Ciencias de la Vida*, *XVII*(31), 237-240.

Molina Pérez, A., & Abkarian, M. (2007, mars). *Cellules minimales et pluralité épistémologique du vivant*. Comunication présenté à II Congrès de la Société de Philosophie des Sciences (SPS), Genève.

Monnard, P. A. (2011, juillet). *Attempt at a systemic design of a protocell: connecting information, metabolism and container*. Comunication présenté à Origins 2011: ISSOL - The International Astronomy Society and Bioastronomy (IAU C51) Joint Conference, Montpellier. Consulté à l'adresse http://www.exobiologie.fr/origins2011/slides/

Monod, J. (1970). *Le hasard et la nécessité: essai sur la philosophie naturelle et la biologie moderne*. Paris: Seuil.

Monod, J., & Jacob, F. (1961). General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 26, p. 389).

Moran, L. A. (2014, juin 25). Sandwalk: The Function Wars: Part I. Consulté 2 janvier 2017, à l'adresse http://sandwalk.blogspot.com.es/2014/06/the-function-wars-part-i.html

Morange, M. (2007). La principale difficulté pour une définition de la vie: concilier continuité et discontinuité. In H. Bersini & J. Reisse (Éd.), *Comment définir la vie?* (p. 65-70). Paris: Vuibert.

Moreira, D., & Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat Rev Micro*, *7*(4), 306-311. https://doi.org/10.1038/nrmicro2108

Moreno, Á., & Mossio, M. (2015). *Biological autonomy: a philosophical and theoretical enquiry*. Consulté à l'adresse http://public.eblib.com/choice/publicfullrecord.aspx?p=2095498

Morowitz, H. J. (1993). *The Beginning of Cellular Life*. Yale University Press.

Morowitz, H. J. (1999). A theory of biochemical organization, metabolic pathways, and evolution. *Complexity*, *4*(6), 39–53.

Morowitz, H. J., & Smith, E. (2007). Energy flow and the organization of life. *Complexity*, *13*(1), 51–59.

Mossio, M., & Bich, L. (2014). What makes biological organisation teleological? *Synthese*, 1-26. https://doi.org/10.1007/s11229-014-0594-z

Mossio, M., Bich, L., & Moreno, A. (2013). Emergence, closure and inter-level causation in biological systems. *Erkenntnis*, *78*(2), 153-178. https://doi.org/10.1007/s10670-013-9507-7

Mossio, M., Montévil, M., & Longo, G. (2016). Theoretical principles for biology: organization. *Progress in Biophysics and Molecular Biology*. https://doi.org/10.1016/j.pbiomolbio.2016

Mossio, M., & Saborido, C. (2016). Functions, organization and etiology. A reply to Artiga and Martinez. *Acta Biotheoretica*, *64*(3), 263-275. https://doi.org/10.1010441-016-9283-2

Mossio, M., Saborido, C., & Moreno, Á. (2009). An Organizational Account of Biological Functions. *British Journal for the Philosophy of Science*, *60*(4), 813-841.

Mossio, M., Saborido, C., & Moreno, Á. (2010). Fonctions: Normativité, téléologie et organisation. In J. Gayon & A. de Ricqlès (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 159-173). Paris: PUF.

Mueller, U. G., Rehner, S. A., & Schultz, T. R. (1998). The Evolution of Agriculture in Ants. *Science*, *281*(5385), 2034-2038. https://doi.org/10.1126/science.281.5385.2034

Nadeau, R. (1999). *Vocabulaire technique et analytique de l'épistémologie* (1. éd). Paris: Presses Univ. de France.

Nagel, E. (1953). Teleological explanation and teleological systems. In H. Feigl & M. Brodbeck (Éd.), *Readings in the Philosophy of Science*. New York: Appleton.

Nagel, E. (1954). Naturalism Reconsidered. *Proceedings and Addresses of the American Philosophical Association*, *28*, 5-17.

Nagel, E. (1961). *The structure of science : problems in the logic of scientific explanation*. Indianapolis: Hackett.

Nagel, E. (1977a). Functional Explanations in Biology. *The Journal of Philosophy*, *74*(5), 280-301.

Nagel, E. (1977b). Goal-Directed Processes in Biology. *The Journal of Philosophy*, *74*(5), 261-279.

Nanay, B. (2010). A Modal Theory of Function. *Journal of Philosophy*, *107*(8), 412-431. https://doi.org/10.5840/jphil2010107834

Nanay, B. (2014). Teleosemantics Without Etiology. *Philosophy of Science*, *81*(5), 798–810.

NASA. (2017). Exoplanet Exploration. Consulté 3 avril 2017, à l'adresse https://exoplanets.nasa.gov/

National Research Council, C. on the L. of O. L. in P. S., Committee on the Origins and Evolution of Life. (2007). *The Limits of Organic Life in Planetary Systems*. Washington, D.C.: The National Academies Press.

Nature. (2007). Meanings of « life ». *Nature*, *447*(7148), 1031-1032. https://doi.org/10.1038/4471031b

Neander, K. (1991a). Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, *58*, 168-184.

Neander, K. (1991b). The Teleological Notion of « Function ». *Australasian Journal of Philosophy*, *69*, 454-468.

Neander, K. (2002). Types of Traits: The Importance of Functional Homologues. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 390-415). Oxford; New York: Oxford University Press.

Neander, K. (2007). Biological Approaches to Mental Representation. In M. Matthen & C. Stephens (Éd.), *Philosophy of Biology*. Elsevier.

Neander, K. (2010). Comment les traits sont-ils typés dans le but de leur attribuer des fonctions? In J. Gayon & A. de Ricqlès (Éd.), *Les Fonctions: des organismes aux artefacts* (p. 99-124). Paris: PUF.

Neander, K. (2012a). Biological function. In E. Craig (Éd.), *Routledge Encyclopedia of Philosophy*. London: Routledge.
Neander, K. (2012b). Teleological Theories of Mental Content. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Spring 2012). Metaphysics Research Lab, Stanford University. Consulté à l'adresse https://plato.stanford.edu/archives/spr2012/entries/content-teleological/
Neander, K., & Rosenberg, A. (2012). Solving the circularity problem for functions. *The Journal of Philosophy*, *109*(10), 613–622.
Nicholson, D. J. (2014). The return of the organism as a fundamental explanatory concept in biology. *Philosophy Compass*, *9*(5), 347-59.
Nissen, L. (1997). *Teleological Language in the Life Sciences*. Oxford: Rowman and Littlefield.
Nunes-Neto, N., Moreno, Á., & El-Hani, C. N. (2014). Function in Ecology: An Organizational Approach. *Biology and Philosophy*, *29*(1), 123–141.
Oberholzer, T., Wick, R., Luisi, P. L., & Biebricher, C. K. (1995). Enzymatic RNA Replication in Self-Reproducing Vesicles: An Approach to a Minimal Cell. *Biochemical and Biophysical Research Communications*, *207*(1), 250-257. https://doi.org/10.1006/bbrc.1995.1180
Oliver, T. H., Mashanova, A., Leather, S. R., Cook, J. M., & Jansen, V. A. A. (2007). Ant semiochemicals limit apterous aphid dispersal. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1629), 3127-3131. https://doi.org/10.1098/rspb.2007.1251
Opfer, J. E., & Gelman, S. A. (2001). Children's and Adults' Models for Predicting Teleological Action: The Development of a Biology-Based Model. *Child Development*, *72*(5), 1367-1381.
Orgel, L. E. (1992). Molecular replication. *Nature*, *358*, 203-209.
Orgel, L. E. (1998). The origins of life - a review of facts and speculations. *TIBS*, *23*, 491-495.
Orgel, L. E. (2004). Prebiotic chemistry and the origin of the RNA world. *Critical reviews in biochemistry and molecular biology*, *39*(2), 99–123.
Origin of Modern Sharks. (2010). Consulté 31 mars 2010, à l'adresse http://www.elasmo-research.org/education/evolution/origin_modern.htm
Orzack, S. H., & Forber, P. (2010, septembre). Adaptationism. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University. Consulté à l'adresse http://plato.stanford.edu/entries/fitness/
Orzack, S. H., & Sober, E. (1994). Optimality models and the test of adaptationism. *The American Naturalist*, *143*(3), 361–380.
Oster, G. F., & Wilson, E. O. (1978). *Caste and Ecology in the Social Insects*. Princeton University Press.
Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Current Biology*, *19*(5), R190-R191. https://doi.org/10.1016/j.cub.2009.01.010
Osvath, M., & Karvonen, E. (2012). Spontaneous Innovation for Future Deception in a Male Chimpanzee. *PLoS ONE*, *7*(5), e36782. https://doi.org/10.1371/journal.pone.0036782
Pacherie, E. (s. d.). *Catégorisation et conceptualisation: homogénéité ou hétérogénéité des processus?*
Palacci, J., Sacanna, S., Steinberg, A. P., Pine, D. J., & Chaikin, P. M. (2013). Living Crystals of Light-Activated Colloidal Surfers. *Science*, *339*(6122), 936-940. https://doi.org/10.1126/science.1230020
Palyi, G., Zucchi, C., & Caglioti, L. (Éd.). (2002). *Fundamentals of Life*. Elsevier.
Papineau, D. (1984). Representation and Explanation. *Philosophy of Science*, *51*(December), 550–72.

Papineau, D. (1987). *Reality and representation*. New York: B. Blackwell.
Papineau, D. (2005). Philosophical Problems of Biology. In T. Honderich (Éd.), *The Oxford Companion to Philosophy* (2e éd.). New York: Oxford University Press.
Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, *348*(6296), 27–33.
Paton, K. R., Varrla, E., Backes, C., Smith, R. J., Khan, U., O'Neill, A., … Coleman, J. N. (2014). Scalable production of large quantities of defect-free few-layer graphene by shear exfoliation in liquids. *Nature Materials*, *13*(6), 624-630. https://doi.org/10.1038/nmat3944
Perlman, M. (2002). Pagan teleology: Adaptational role and the philosophy of mind. In *Functions: New Essays in the Philosophy of Psychology and Biology* (p. 263-90). Oxford: Oxford Univ Pr. Consulté à l'adresse http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=phl&AN=PHL1703998&site=ehost-live&scope=site
Perlman, M. (2004). The Modern Philosophical Resurrection of Teleology. *Monist: An International Quarterly Journal of General Philosophical Inquiry*, *87*(1), 3-51.
Perlman, M. (2009). Changing the mission of theories of teleology: DOs and DON'Ts for thinking about function. In U. Krohs & P. Kroes (Éd.), *Functions in Biological and Artificial Worlds* (p. 17-36). Cambridge, MA; London: MIT Press.
Perret, N. (2015). Téléologie biologique aujourd'hui, entre transcendantal et naturalisme. *Studi Kantiani*, *28*, 89–104.
Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M., … Abergel, C. (2013). Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science*, *341*(6143), 281-286. https://doi.org/10.1126/science.1239181
Piaget, J. (1947). *La représentation du monde chez l'enfant*. Paris: PUF.
Pichot, A. (1993). *Histoire de la notion de vie*. Paris: Gallimard.
Pigliucci, M., & Kaplan, J. (2000). The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. *Trends in Ecology & Evolution*, *15*(2), 66–70.
Pinker, S. (2000). *Comment fonctionne l'esprit*. Paris: Odile Jacob.
Pirie, N. W. (1957). The Origins of Life: Moscow Symposium. *Nature*, *180*(4592), 886-888. https://doi.org/10.1038/180886a0
Plantinga, A. (1993). *Warrant and Proper Function*. Oxford: Oxford University Press.
Pohorille, A. (2011, juillet). *Origins of protein functions in cells*. Comunication présenté à Origins 2011: ISSOL - The International Astronomy Society and Bioastronomy (IAU C51) Joint Conference, Montpellier. Consulté à l'adresse http://www.exobiologie.fr/origins2011/slides/
Popa, R. (2004). *Between Necessity and Probability: Searching for the Definition and Origin of Life* (2004 edition). Berlin ; New York: Springer.
Popa, R. (2010). Necessity, futility and the possibility of defining life are all embedded in its origins as a punctuated-gradualism. *Origins of Life and Evolution of Biospheres*, (40), 183-190.
Popper, K. (2005). *Unended Quest: An Intellectual Autobiography*. Routledge.
Premack, D., & Premack, A. J. (1997). Infants attribute value to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, *9*(6), 848–856.
Preston, B. (1998). Why is a Wing Like a Spoon? A Pluralist Theory of Function. *The Journal of Philosophy*, *95*(5), 215-254.
Price, C. (1995). Functional Explanations and Natural Norms. *Ratio: An International Journal of Analytic Philosophy*, *8*(2), 143-160.

Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P., … others. (2012). Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science*, *335*(6070), 843–847.

Pross, A. (2011). Toward a general theory of evolution: Extending Darwinian theory to inanimate matter. *Journal of Systems Chemistry*, *2*(1), 1. https://doi.org/10.1186/1759-2208-2-1

Proust, J. (1995). Fonction et causalité. *Intellectica*, *21*(2), 81-113.

Pruetz, J. D., Bertolani, P., Ontl, K. B., Lindshield, S., Shelley, M., & Wessling, E. G. (2015). New evidence on the tool-assisted hunting exhibited by chimpanzees (*Pan troglodytes verus*) in a savannah habitat at Fongoli, Sénégal. *Royal Society Open Science*, *2*(4), 140507. https://doi.org/10.1098/rsos.140507

Purton, A. C. (1979). Biological function. *The Philosophical Quarterly*, *29*(114), 10-24.

Pyke, G. H. (1984). Optimal foraging theory: a critical review. *Annual review of ecology and systematics*, 523–575.

Quine, W. V. O. (1960). *Word and object*. Cambridge, Mass: MIT Press.

Radman, M. (2002). *Mutation, évolution et sélection* [Streaming]. Consulté à l'adresse http://www.canal-u.tv/producteurs/universite_de_tous_les_savoirs/dossier_programmes/les_conferences_de_l_annee_2002/la_diversite_de_la_vie/mutation_evolution_et_selection

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., … Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science*, *306*(5700), 1344.

Raoult, D., & Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. *Nat Rev Micro*, *6*(4), 315-319. https://doi.org/10.1038/nrmicro1858

Rasmussen, S., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., Stadler, P. F., & Bedau, M. A. (2004). Transitions from Nonliving to Living Matter. *Science*, *303*.

Rasmussen, S., Chen, L., Nilsson, M., & Abe, S. (2003). Bridging Nonliving and Living Matter. *Artificial Life*, *9*(3), 269-317.

Ratcliffe, M. (2000). The function of function. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 113-133. https://doi.org/10.1016/S1369-8486(99)00039-4

Ratcliffe, M. (2001). A Kantian Stance on the Intentional Stance. *Biology and Philosophy*, *16*(1), 29-52.

Ratcliffe, M. (2003). Teleology and the asumption of naturalism. *Metascience*, *12*(3), 312-321. https://doi.org/10.1023/B:MESC.0000005817.10723.84

Reeve, H. K., & Sherman, P. W. (1993). Adaptation and the goals of evolutionary success. *The Quarterly Review of Biology*, *68*(1), 1-32.

Reisser, J., Shaw, J., Hallegraeff, G., Proietti, M., Barnes, D. K. A., Thums, M., … Pattiaratchi, C. (2014). Millimeter-Sized Marine Plastics: A New Pelagic Habitat for Microorganisms and Invertebrates. *PLoS ONE*, *9*(6), e100289. https://doi.org/10.1371/journal.pone.0100289

Ricardo, A., & Szostak, J. W. (2009, novembre). El orígen de la vida. *Investigación y Ciencia*, *398*, 38-46.

Ridley, M. (1998). Principles of classification. In M. Ruse (Éd.), *The Philosophy of Biology* (p. 167-179). Prometheus Books.

Ritzenberg, A. L., Adam, D. R., & Cohen, R. J. (1984). Period multupling-evidence for nonlinear behaviour of the canine heart. *Nature*, *307*(5947), 159-161. https://doi.org/10.1038/307159a0

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anat Embryol*, *210*, 419-421.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141.
Röhl, J., & Jansen, L. (2014). Why functions are not special dispositions: an improved classification of realizables for top-level ontologies. *Journal of Biomedical Semantics*, *5*(1), 27. https://doi.org/10.1186/2041-1480-5-27
Rosander, K., & von Hofsten, C. (2004). Infants' emerging ability to represent occluded object motion. *Cognition*, *91*(1), 1-22.
Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, *4*, 328-350.
Rosch, E. (1975). Cognitive Reference Points. *Cognitive Psychology*, *7*, 532-547.
Rosch, E. (1978). Principles of Categorization. In *Cognition and Categorization* (p. 27-48). Hillsdale, N. J.: Lawrence Erlbaum.
Rose, M. R., & Lauder, G. V. (1996). Post-spandrel adaptationism. In M. R. Rose & G. V. Lauder (Éd.), *Adaptation* (p. 1–8). Academic Press.
Rosen, R. (2013). *Optimality principles in biology*. Springer.
Rosenberg, A. (1983). Fitness. *The Journal of Philosophy*, *80*(8), 457–473.
Rosenberg, A. (1985). *The Structure of Biological Science*. Cambridge: Cambridge University Press.
Rosenberg, A. (2001a). How is Biological Explanation Possible? *52*, *4*, 735-760.
Rosenberg, A. (2001b). Reductionism in a historical science. *Philosophy of Science*, *68*(2), 135–163.
Rosenberg, A., & Bouchard, F. (2010, septembre). Fitness. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University.
Rosenberg, A., & Neander, K. (2009). Are Homologies (Selected Effect or Causal Role) Function Free? *Philosophy of Science*, *76*(3), 307-334.
Rosenblueth, A., & Wiener, N. (1950). Purposeful and non-purposeful behavior. *Philosophy of Science*, *17*(4), 318–326.
Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of science*, *10*(1), 18–24.
Rowe, S. J. (1992). Biological Fallacy: Life Equals Organisms. *BioScience*, *42*(6), 394.
Rowe, S. J. (1998). « Earth » as the Metaphor for « Life ». *BioScience*, *48*(6), 428-429.
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science*, *26*(5), 521-562. https://doi.org/http://dx.doi.org/
Ru, V. L. (1994). *Jean Le Rond d'Alembert philosophe*. Vrin.
Ruffman, T., Slade, L., & Redman, J. (2005). Young infants' expectations about hidden objects. *Cognition*, *97*(2), B35-B43.
Ruiz, A. M., & Santos, L. R. (2013). Understanding differences in the way human and non-human primates represent tools: The role of teleological-intentional information. *Tool Use in Animals: Cognition and Ecology*, 119.
Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2010). Defining life or bringing biology to life. *Origins of Life and Evolution of Biospheres*, (40), 203-13.
Runyon, K. D., Stern, S. A., Lauer, T. R., Grundy, W., Summers, M. E., & Singer, K. N. (2017). A geophysical planet definition. Présenté à 48th Lunar and Planetary Science Conference, held 20-24 March 2017, The Woodlands, Texas.
Ruse, M. (1971). Functional Statements in Biology. *Philosophy of Science*, *38*(1), 87-95.
Ruse, M. (1973a). A Reply to Wright's Analysis of Functional Statements. *Philosophy of Science*, *40*(2), 277-280.
Ruse, M. (1973b). *The philosophy of biology*. London: Hutchinson.

Ruse, M. (1981). The Last Word on Teleology, or Optimality Models Vindicated. In *Is Science Sexist?* (p. 85-101). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-8443-1_4
Ruse, M. (1982). Teleology Redux. In J. Agassi & R. S. Cohen (Éd.), *Scientific Philosophy Today: Essays in Honor of Mario Bunge* (p. 299-309). Dordrecht; Boston: Reidel (Boston Studies in the Philosophy of Science 67).
Ruse, M. (2002). Evolutionary biology and teleological thinking. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 33-59). Oxford; New York: Oxford University Press.
Russell, J. R., Huang, J., Anand, P., Kucera, K., Sandoval, A. G., Dantzler, K. W., … Strobel, S. A. (2011). Biodegradation of Polyester Polyurethane by Endophytic Fungi. *Applied and Environmental Microbiology*, *77*(17), 6076-6084. https://doi.org/10.1128/AEM.00521-11
Saborido, C. (2012). *Funcionalidad y organización en biología. Reformulación del concepto de función biológica desde una perspectiva organizacional.* Universidad del País Vasco, San Sebastian.
Saborido, C., Mossio, M., & Moreno, Á. (2011). Biological Organization and Cross-Generation Functions. *British Journal for the Philosophy of Science*, *62*(3), 583-606.
Sagan, C. (1970). Life. In *Encyclopedia Britannica*.
Saïb, A. (2006, décembre). Les virus, inertes ou vivants? *Pour la Science*, *350*, 60-64.
Salva, O. R., Regolin, L., & Vallortigara, G. (2012). Inversion of contrast polarity abolishes spontaneous preferences for face-like stimuli in newborn chicks. *Behavioural brain research*, *228*(1), 133–143.
Santos, L. R., Miller, C. T., & Hauser, M. D. (2003). Representing tools: how two non-human primate species distinguish between the functionally relevant and irrelevant features of a tool. *Animal cognition*, *6*(4), 269–281.
Sarkar, S. (2000). Information in Genetics and Developmental Biology: Comments on Maynard Smith. *Philosophy of Science*, *67*(2), 208-213.
Scheffler, I. (1959). Thoughts on Teleology. *The British Journal for the Philosophy of Science*, *9*(36), 265-284.
Schlosser, G. (1998). Self-Re-Production and Functionality: A Systems-Theoretical Approach to Teleological Explanation. *Synthese*, *116*(3), 303-354.
Schlosser, G. (2003). Naturalizing Functions Unity Beyond Pluralism? *Studies in History and Philosophy of Biological and Biomedical Sciences*, *34C*, 697.
Schrödinger, E. (1944). *Qu'est-ce que la vie? De la physique à la biologie*. Paris: Seuil.
Schwartz, P. H. (2002). The Continuing Usefulness Account of Proper Function. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 244-269). Oxford; New York: Oxford University Press.
Schwartz, P. H. (2004). An Alternative to Conceptual Analysis in the Function Debate. *Monist: An International Quarterly Journal of General Philosophical Inquiry*, *87*(1), 136-153.
ScienceAlert. (2014, septembre 30). Pluto's a Planet, According to a New Public Debate. Consulté 10 avril 2017, à l'adresse http://www.sciencealert.com/plutos-a-planet-according-to-a-new-public-debate-that-questioned-definition-of-planet
Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge Mass.: MIT Press.
Searle, J. R. (1995). *The Construction of Social Reality*. The Free Press.
Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-Month-Olds Really Attribute Mental States to Others? *Psychological science*, *22*(7), 878–880.
Serrano, L. (2007). Synthetic biology: promises and challenges. *Mol Syst Biol*, *3*. https://doi.org/10.1038/msb4100202

Shapiro, R. (2007). A simpler origin of life. *Scientific American*, *June 2007*, 24-31.
Shetty, R., Endy, D., & Knight, T. (2008). Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, *2*(1), 5. https://doi.org/10.1186/1754-1611-2-5
Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. JHU Press.
Siegel, E. (2015, janvier 7). How many habitable planets are in our galaxy? Consulté 27 août 2015, à l'adresse https://medium.com/starts-with-a-bang/how-many-habitable-planets-are-in-our-galaxy-5bcf6db80c7f
Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809-813. https://doi.org/10.1073/pnas.0707021105
Simon, T. W. (1976). A Cybernetic Analysis of Goal-Directedness. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1976*, 56-67.
Sinigaglia, C. (2008). 2 Enactive Understanding and Motor Intentionality. In F. Morganti, A. Carassa, & G. Riva (Éd.), *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*. Amsterdam: IOS Press.
Sloman, S. A., & Fernbach, P. M. (2008). The value of rational analysis: An assessment of causal reasoning and learning. *The probabilistic mind: Prospects for rational models of cognition*, 485–500.
Soavi, M. (2009). Antirealism and Artefact Kinds. *Techne*, *13*(2), 93–107.
Sober, E. (1984). *The Nature of Selection*. Cambridge, MA: MIT Press.
Sober, E. (1993). *Philosophy of Biology*. Boulder: Westview Press.
Sober, E. (2001). The two faces of fitness. *Thinking about evolution: historical, philosophical, and political perspectives*, *2*, 309–321.
Sommerhoff, G. (1950). *Analytical biology*. Oxford University Press.
Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-B11. https://doi.org/10.1016/j.cognition.2004.07.004
Sorabji, R. (1964). Function. *The Philosophical Quarterly*, *14*(57), 289-302.
Sorensen, R. (2008). Vagueness. In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Consulté à l'adresse http://plato.stanford.edu/entries/vagueness/
Soter, S. (2007). What is a planet? *Scientific American Magazine*, *296*(1), 34–41.
Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059-1069. https://doi.org/10.1016/j.cognition.2007.10.002
Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, *21*(3), 355–359.
Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, *5*(6), 769–772.
Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.
Sperber, D. (2007). Seedless grapes: nature and culture. In E. Margolis & S. Laurence (Éd.), *Creations of the Mind: Theories of Artifacts and Their Representation* (p. 124-137). Oxford University Press.
Sprinzak, D., & Elowitz, M. B. (2005). Reconstruction of genetic circuits. *Nature*, *438*(7067), 443-448. https://doi.org/10.1038/nature04335

Srinivasan, V., & Morowitz, H. J. (2009a). Analysis of the Intermediary Metabolism of a Reductive Chemoautotroph. *The Biological Bulletin*, *217*(3), 222.

Srinivasan, V., & Morowitz, H. J. (2009b). The canonical network of autotrophic intermediary metabolism: minimal metabolome of a reductive chemoautotroph. *The Biological Bulletin*, *216*(2), 126.

St Amant, R., & Horton, T. E. (2008). Revisiting the definition of animal tool use. *Animal Behaviour*, *75*(4), 1199-1208. https://doi.org/10.1016/j.anbehav.2007.09.028

Stegmann, U. E. (2005). Genetic Information as Instructional Content. *Philosophy of Science*, *72*, 425-443.

Stevens, M., Yule, D. H., & Ruxton, G. D. (2008). Dazzle coloration and prey movement. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1651), 2639-2643. https://doi.org/10.1098/rspb.2008.0877

Strevens, M. (2004). The causal and unification approaches to explanation unified—causally. *Noûs*, *38*(1), 154–176.

Sumpter, D. J., & Beekman, M. (2003). From nonlinearity to optimality: pheromone trail foraging by ants. *Animal behaviour*, *66*(2), 273–280.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580.

Szathmary, E. (2005). In search of the simplest cell. *Nature*, *433*, 469-470.

Szostak, J. W. (2008, juin 4). The Origins of Function in Biological Nucleic Acids, Proteins, and Membranes. Consulté 21 février 2009, à l'adresse http://www.hhmi.org/research/investigators/szostak.html

Szostak, J. W. (2012). Attempts to define life do not help to understand the origin of life. *Journal of Biomolecular Structure & Dynamics*, *29*(4), 599-600. https://doi.org/10.1080/073911012010524998

Szostak, J. W., Bartel, D. P., & Luisi, P. L. (2001). Synthetizing life. *Nature*, *409*, 387-390.

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, *349*(6249), 734-738. https://doi.org/10.1126/science.aab1058

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *PNAS*, *104*(46), 17954-17958.

Taylor, R. (1950a). Comments on a Mechanistic Conception of Purposefulness. *Philosophy of Science*, *17*(4), 310-317.

Taylor, R. (1950b). Purposeful and Non-Purposeful Behavior: A Rejoinder. *Philosophy of Science*, *17*(4), 327-332.

Terada, K., Shamoto, T., & Ito, A. (2008). Human goal attribution toward behavior of artifacts. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on* (p. 160–165). Consulté à l'adresse http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4600660

Terada, K., Shamoto, T., Mei, H., & Ito, A. (2007). Reactive movements of non-humanoid robots cause intention attribution in humans. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on* (p. 3715–3720). Consulté à l'adresse http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4399429

Toepfer, G. (2012). Teleology and Its Constitutive Role for Biology as the Science of Organized Systems in Nature. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 113-119.

Trent, J. D. (2007). Extremophiles in Astrobiology: PerArdua ad Astra. *Gravitational and Space Biology*, *13*(2).
Trifonov, E. N. (2011). Vocabulary of definitions of life suggests a definition. *Journal of Biomolecular Structure and Dynamics*, *29*(2), 259–266.
Trifonov, E. N. (2012). Definition of life: Navigation through Uncertainties. *Journal of Biomolecular Structure and Dynamics*, *29*(4), 647-650.
Uexküll, J. von. (1956). *Mondes animaux et monde humain*. Paris: Denoël.
Vaesen, K., & van Amerongen, M. (2008). Optimality vs. Intent: Limitations of Dennett's Artifact Hermeneutics. *Philosophical Psychology*, *21*(6), 779–797.
Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually Inexperienced Chicks Exhibit Spontaneous Preference for Biological Motion Patterns. *PLoS Biol*, *3*(7), e208. https://doi.org/10.1371/journal.pbio.0030208
Varela, F. J., Maturana, H., & Uribe, R. (1974). Autopoïesis: the organization of living systems, its caracterization and a model. *BioSystems*, *5*, 187-195.
Vásconez, M., & Peña, L. (1996). ¿Qué es una ontología gradual? *Agora*, *15*(2), 29-48.
Vega Encabo, J. (2009). Estado de la cuestión: Filosofía de la tecnología. *Theoria: Revista de Teoria, Historia y Fundamentos de la Ciencia*, *24:3*(66), 323-341.
Vermaas, P. E., Carrara, M., Borgo, S., & Garbacz, P. (2013). The design stance and its artefacts. *Synthese*, *190*(6), 1131–1152.
Villarreal, L. P., & Witzany, G. (2010). Viruses are essential agents within the roots and stem of the tree of life. *Journal of Theoretical Biology*, *262*(4), 698-710. https://doi.org/10.1016/j.jtbi.2009.10.014
von Wright, G. H. (1963). *The Varieties of Goodness*. Routledge and Kegan Paul.
Vreeland, R. H., Rosenzweig, W. D., & Powers, D. W. (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature*, *407*(6806), 897-900. https://doi.org/10.1038/35038060
Wachbroit, R. (1994). Normality as a Biological Concept. *Philosophy of Science*, *61*(4), 579-591.
Wade, N. (2012, mai 30). The Tomato: Ripe, Juicy and Bursting With Genes. *The New York Times*. Consulté à l'adresse http://www.nytimes.com/2012/05/31/science/the-tomato-ripe-juicy-and-bursting-with-genes.html
Walde, P., Wick, R., Fresta, M., Mangone, A., & Luisi, P. L. (1994). Autopoietic self-reproduction of fatty acid vesicles. *Journal of the American Chemical Society*, *116*(26), 11649–11654.
Walsh, D. M. (1996). Fitness and Function. *British Journal for the Philosophy of Science*, *47*(4), 553-574.
Walsh, D. M. (2000). Chasing shadows: natural selection and adaptation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *31*(1), 135-153. https://doi.org/10.1016/S1369-8486(99)00041-2
Walsh, D. M. (2002). Brentano's Chestnuts. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 314-337). Oxford; New York: Oxford University Press.
Walsh, D. M., & Ariew, A. (1996). A Taxonomy of Functions. *Canadian Journal of Philosophy*, *26*(4), 493-514.
Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences*, *1*(2), 97–125.

Westall, F., Foucher, F., Bost, N., Bertrand, M., Loizeau, D., Vago, J. L., … Cockell, C. S. (2015). Biosignatures on Mars: What, Where, and How? Implications for the Search for Martian Life. *Astrobiology*, *15*(11), 998-1029. https://doi.org/10.1089/ast.2015.1374

Wiener, N. (1948). *Cybernetics*. New York: J. Wiley.

Willatts, P. (1999). Development of means–end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental psychology*, *35*(3), 651.

Williams, G. C. (1966). *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton Univ Pr.

Wimsatt, W. (1972). Teleology and the Logical Structure of Function Statement. *Studies in the History and Philosophy of Science*, *3*(1), 1-80.

Wimsatt, W. (2002). Functional Organisation, Analogy, and Inference. In A. Ariew, R. Cummins, & M. Perlman (Éd.), *Functions: new essays in the philosophy of psychology and biology* (p. 63-112). Oxford; New York: Oxford University Press.

Wittgenstein, L. (1984). *Philosophical investigations*. (G. E. M. Anscombe, Trad.) (3e éd.). Oxford: B. Blackwell.

Wolfe-Simon, F., Blum, J. S., Kulp, T. R., Gordon, G. W., Hoeft, S. E., Pett-Ridge, J., … Oremland, R. S. (2010). A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science*. https://doi.org/10.1126/science.1197258

Woodfield, A. (1976). *Teleology*. New York: Cambridge University Press.

Woodfield, A. (1998). Teleology. In *Routledge Encyclopedia of Philosophy*. London: Taylor & Francis. Consulté à l'adresse https://www.rep.routledge.com/articles/thematic/teleology/v-1

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34. https://doi.org/10.1016/S0010-0277(98)00058-4

Woodward, J. (1997). Explanation, Invariance, and Intervention. *Philosophy of Science*, *64*, 26-46.

Woodward, J. (2000). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science*, *51*(2), 197-254.

Woodward, J. (2001). Law and Explanation in Biology: Invariance Is the Kind of Stability That Matters. *Philosophy of Science*, *68*(1), 1-20.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318.

Wouters, A. (2005). The Function Debate in Philosophy. *Acta Biotheoretica*, *53*(2), 123-151. https://doi.org/10.1007/s10441-005-5353-6

Wright, L. (1968). The Case against Teleological Reductionism. *The British Journal for the Philosophy of Science*, *19*(3), 211-223.

Wright, L. (1972). Explanation and Teleology. *Philosophy of Science*, *39*(2), 204-218.

Wright, L. (1973). Functions. *The Philosophical Review*, *82*(2), 139-168.

Wright, L. (1976). *Teleological explanations: an etiological analysis of goals and functions*. Los Angeles: California University Press.

Wright, L. (2013). Epilogue. In P. Huneman (Éd.), *Functions: Selection and Mechanisms* (p. 233-43). Springer.

Wunenburger, J.-J. (2000). Métaphore, poiétique et pensée scientifique. *Revue européenne des sciences sociales. European Journal of Social Sciences*, (XXXVIII-117), 35-47. https://doi.org/10.4000/ress.707

Xu, W., Edwards, M. R., Borek, D. M., Feagins, A. R., Mittal, A., Alinger, J. B., … Amarasinghe, G. K. (2014). Ebola virus VP24 targets a unique NLS binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated STAT1. *Cell Host & Microbe*, *16*(2), 187-200. https://doi.org/10.1016/j.chom.2014.07.008

Yang, J., Cowan, N. B., & Abbot, D. S. (2013). Stabilizing Cloud Feedback Dramatically Expands the Habitable Zone of Tidally Locked Planets. *The Astrophysical Journal Letters*, *771*(2), L45. https://doi.org/10.1088/2041-8205/771/2/L45

Yang, J., Yang, Y., Wu, W.-M., Zhao, J., & Jiang, L. (2014). Evidence of polyethylene biodegradation by bacterial strains from the guts of plastic-eating waxworms. *Environmental science & technology*, *48*(23), 13776–13784.

Yong, E. (2013). Giant viruses open Pandora's box. *Nature*. https://doi.org/10.1038/nature.2013.13410

Zach, R. (1978). Selection and dropping of whelks by northwestern crows. *Behaviour*, *67*(1), 134–147.

Zellner, H. M. (2001). Wright's functions and Kitcher's gas. *Philosophia*, *28*(1), 503–509.

Zepik, H. H., Bloechliger, E., & Luisi, P. L. (2001). A chemical model of homeostasis. *Angew. Chem. Int. Ed. Engl.*, *40*, 199-202.

INDEX DES ILLUSTRATIONS

TABLE DES MATIÈRES

Troisième partie

CARACTÉRISATION GÉNÉRALE DES FONCTIONS

Quatrième partie

JUSTIFICATION DE LA TÉLÉOLOGIE